



UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Dados abertos na administração pública de cidades inteligentes promovendo transparência aos cidadãos

Dissertação de Mestrado

Vinícius Almeida Teles Barreto



São Cristóvão – Sergipe

2019

UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Vinícius Almeida Teles Barreto

**Dados abertos na administração pública de cidades
inteligentes promovendo transparência aos cidadãos**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Sergipe como requisito parcial para a obtenção do título de mestre em Ciência da Computação.

Orientador(a): Prof. Dr. Rogério Patrício Chagas do Nascimento

Coorientador(a): Prof. Dr. Michel dos Santos Soares

São Cristóvão – Sergipe

2019

**FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL
UNIVERSIDADE FEDERAL DE SERGIPE**

B273 Barreto, Vinícius Almeida Teles
Dados abertos na Administração pública de cidades inteligentes promovendo transparência aos cidadãos / Vinícius Almeida Teles Barreto ; orientador Rogério Patrício Chagas do Nascimento . - São Cristóvão, 2019.
88 f. : il.

Dissertação (mestrado em Ciência da Computação) – Universidade Federal de Sergipe, 2019.

1. Computação. 2. Serviços de informação. 3. Informação pública. 4. Transparência na administração pública. 5. Cidades inteligentes. 6. Internet das coisas. I. Nascimento, Rogério Patrício Chagas do orient. II. Título.

CDU 004:35



UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
COORDENAÇÃO DE PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**Ata da Sessão Solene de Defesa da Dissertação do
Curso de Mestrado em Ciência da Computação-UFS.
Candidato: VINICIUS ALMEIDA TELES BARRETO**

Em 24 dias do mês de janeiro do ano de dois mil e dezenove, com início às 14h00min, realizou-se na Sala de Seminários do DCOMP da Universidade Federal de Sergipe, na Cidade Universitária Prof. José Aloísio de Campos, a Sessão Pública de Defesa de Dissertação de Mestrado do candidato **Vinicius Almeida Teles Barreto**, que desenvolveu o trabalho intitulado: **"Dados abertos na administração pública de cidades inteligentes promovendo transparência aos cidadãos"**, sob a orientação do Prof. Dr. **Rogério Patrício Chagas do Nascimento**. A Sessão foi presidida pelo Prof. Dr. **Rogério Patrício Chagas do Nascimento** (PROCC/UFS), que após a apresentação da dissertação passou a palavra aos outros membros da Banca Examinadora, Prof. Dr. **Michel dos Santos Soares** (PROCC/UFS) e, em seguida, ao Prof. **Jefferson David Araujo Sales** (UFS). Após as discussões, a Banca Examinadora reuniu-se e considerou o mestrando (a) APROVADO "(aprovado/reprovado)" _____ "(com/sem)" _____ ressalvas. Atendidas as exigências da Instrução Normativa 01/2017/PROCC, do Regimento Interno do PROCC (Resolução 67/2014/CQNEPE), e da Resolução nº 25/2014/CONEPE que regulamentam a Apresentação e Defesa de Dissertação, e nada mais havendo a tratar, a Banca Examinadora elaborou esta Ata que será assinada pelos seus membros e pelo mestrando.

Cidade Universitária "Prof. José Aloísio de Campos", 24 de Janeiro de 2019.

Prof. Dr. Rogério Patrício Chagas do
Nascimento
(PROCC/UFS)
Presidente

Prof. Dr. Michel dos Santos Soares
(PROCC/UFS)
Examinador Interno

Prof. Dr. Jefferson David Araujo Sales
(UFS)
Examinador Externo

Vinicius Almeida Teles Barreto
Candidato

Resumo

A proposta de uma cidade inteligente é gerir de maneira eficaz os recursos de uma população, fazendo uso de informações geradas e coletadas no dia-a-dia dos cidadãos por meio de Tecnologias de Informação e Comunicação (TIC), e Internet das Coisas, do inglês *Internet of Things* (IoT). Assim, os dados das cidades podem ser armazenados, tratados e publicados de maneira que fiquem acessíveis à população, promovendo transparência no uso de bens públicos. Dessa forma, os desenvolvedores podem utilizar os dados em sistemas de informação e torná-los aliados dos cidadãos no provimento de bem-estar, divulgação de conhecimento, combate à poluição do meio-ambiente, estímulo da economia, combate à corrupção, dentre outras vantagens para a cidade. No Brasil, a criação da lei nº 12.527/2011, popularmente conhecida como Lei de Acesso à Informação (LAI), garante o direito constitucional de acesso às informações governamentais por qualquer pessoa física ou jurídica, sem a necessidade de informar motivos de solicitação. Com o objetivo de promover a transparência em cidades inteligentes, este trabalho propõe uma arquitetura de publicação de dados abertos governamentais com base na *ISO/IEC/IEEE 42010:2011* que resultou no desenvolvimento de um Plano de Dados Abertos (PDA), e por meio dele, o desenvolvimento do Portal de Dados Abertos da UFS com um conjunto inicial de dados. Por fim, três bibliotecas de código aberto, que combinadas resultam numa solução para publicação de dados abertos, foram desenvolvidas para instituições de ensino público superior a fim de auxiliar o processo de extração e publicação das informações.

Palavras-chave: Dados Abertos, Plano de Dados Abertos, Arquitetura de Dados Abertos, Dados Abertos Governamentais, Transparência, Cidades Inteligentes, Internet das Coisas.

Abstract

The smart cities purposes is a efficient management of population resources, using daily generated and collected citizens data through Information and Communication Technologies (*ICT*) and Internet of Things (*IoT*). Furthermore, the city's data can be stored, processed and published being accessible to the public, promoting transparency in assets use. Hence IT developers can use this data on systems and turn them, supporting citizens on welfare, knowledge sharing, fighting air pollution, economic activities raising, anti-corruption tasks, as well as others benefits. In Brazil, the law nº 12,527/2011, commonly as known as Information Access Law (*LAI*, in brazilian portuguese initials), ensures the constitutional right for any individual or legal entity to access government information with no need of indicate requirement reasons. Setting the transparency promotion as its main goal, this work proposes an open government data architecture based on *ISO/IEC/IEEE 42010:2011* resulting on an Open Data Plan (*ODP*), and through it, it was possible develop the Open Data UFS Portal, with an initial number of datasets. At last, three open-source libraries, which combined result in a solution for open data publishing, were developed for higher education institutions to assist in the data extraction and publication.

Keywords: Open Data, Open Data Plan, Open Data Architecture, Open Government Data, Transparency, Smart Cities, Internet of Things (IoT).

Lista de ilustrações

Figura 1 – Portal Transparência Traduzida	20
Figura 2 – Escada de evolução no modelo 5 estrelas de publicação de dados abertos, definido por Tim Berners-Lee	21
Figura 3 – Portal Alagoas em Dados e Informações	23
Figura 4 – Portal Copenhagen Data	23
Figura 5 – Contexto da Descrição de uma Arquitetura	25
Figura 6 – Modelo Conceitual da Descrição de uma Arquitetura	26
Figura 7 – Framework de uma Arquitetura	28
Figura 8 – Simbologia utilizada no diagrama de componentes	29
Figura 9 – Visão Geral da Arquitetura <i>Open Data UFS</i>	33
Figura 10 – Visão Geral do Procedimento de Carga dos Dados no Portal	45
Figura 11 – As Duas Interfaces Necessárias	45
Figura 12 – A Montagem dos Componentes	47
Figura 13 – Representação das Interfaces Necessárias como Componentes	47
Figura 14 – Detalhes do Componente <i>Open Data Link</i>	48
Figura 15 – Visão geral do portal <i>Dados Abertos UFS</i>	60
Figura 16 – Tela de lista de arquivos de um <i>dataset</i>	60
Figura 17 – Tela de pré-visualização e download de um <i>dataset</i>	61

Lista de abreviaturas e siglas

UFS	Universidade Federal de Sergipe
UFRN	Universidade Federal do Rio Grande do Norte
NTI	Núcleo de Tecnologia da Informação da UFS
AUDINT	Auditoria Interna da UFS
CTIn	Comissão de Tecnologia da Informação
IoT	<i>Internet of Things</i>
OKF	<i>Open Knowledge Foundation</i>
www	<i>World Wide Web</i>
REST	<i>Representational State Transfer</i>
URI	<i>Universal Resource Identifier</i>
URL	<i>Universal Resource Location</i>
API	<i>Application Programming Interface</i>
TOGAF	<i>The Open Group Architecture Framework</i>
CKAN	<i>Comprehensive Knowledge Archive Network</i>

Sumário

1	Introdução	10
1.1	Justificativa	11
1.2	Objetivo Geral	12
1.3	Objetivos Específicos	12
1.4	Hipóteses	13
1.5	Metodologia de Pesquisa	13
1.6	Trabalhos Relacionados	14
1.7	Organização da Dissertação	15
2	Fundamentação Teórica	17
2.1	Transparência Governamental	17
2.2	Cidades Inteligentes e Internet das Coisas (<i>IoT</i>)	18
2.3	Dados Abertos	19
2.4	<i>CKAN</i>	22
2.5	Norma ISO/IEC/IEEE 42010:2011	25
2.5.1	Contexto e Modelo Conceitual da Descrição de uma Arquitetura	25
2.5.2	Frameworks e Linguagem de Descrição de uma Arquitetura	28
2.6	Diagrama de Componentes	28
2.7	Padrão Arquitetural REST (Representational State Transfer)	30
2.8	Considerações Finais do Capítulo	31
3	Arquitetura Open Data UFS	32
3.1	Visão Geral	32
3.2	Identificação das Partes Interessadas (<i>Stakeholders</i>)	33
3.3	Interesses (<i>Concerns</i>)	35
3.4	Requisitos funcionais e não-funcionais	36
3.5	Visões e Pontos de Vista (<i>Views e Viewpoints</i>)	37
3.5.1	Ponto de Vista do Negócio	38
3.5.2	Ponto de Vista dos Dados	39
3.5.3	Ponto de Vista da Aplicação	40
3.5.4	Ponto de Vista da Tecnologia	41
3.6	Tecnologias Envolvidas na Implantação do Portal Dados Abertos UFS	41
3.7	Personalização do CKAN para o Portal Dados Abertos UFS	42
3.8	Instalação do Portal Dados Abertos UFS no Servidor	44
3.9	Representação em Componentes do Procedimento de Publicação das Informações no Portal	44

3.9.1	As Duas Interfaces Necessárias	45
3.9.2	A Montagem dos Componentes	46
3.9.3	Representação das Interfaces Necessárias como Componentes	46
3.9.4	Detalhes do Componente Open Data Link	47
3.10	Implementação da Solução Open Data Link	48
3.11	Recursos da API Sistemas	50
3.12	Desafios Superados e Sugestões	51
3.13	Considerações Finais do Capítulo	51
4	Plano de Dados Abertos (PDA) e Portal Dados Abertos UFS	52
4.1	Embasamento Legal	52
4.2	Motivação Social	53
4.3	Alinhamento ao Planejamento Estratégico da Instituição	54
4.4	Alinhamento ao Planejamento Tecnológico da Instituição	55
4.5	Desenvolvimento do PDA e Execução	56
4.6	Aprovação do PDA	57
4.7	Desafios Superados e Sugestões para o Próximo PDA	58
4.8	Apresentação do Portal de Dados Abertos da UFS	59
4.9	Considerações Finais do Capítulo	61
5	Conclusão	62
5.1	Principais Contribuições	62
5.2	Limitações da Pesquisa	63
5.3	Trabalhos futuros	63
	Referências	65
	 Apêndices	 68
	APÊNDICE A Plano de Dados Abertos (PDA) da UFS	69

1

Introdução

O cenário político que o Brasil está imerso incita cada vez mais a participação popular no acompanhamento das ações governamentais e, numa rápida visita aos portais de transparência disponíveis, pode-se perceber que as informações ali publicadas estão recheadas de linguajar incomum à maioria da população. De acordo com [Andersen \(2009\)](#), as ferramentas de governo eletrônico tornaram possível a significativa redução dos casos de corrupção, no período compreendido entre os anos 1996 e 2006. É relevante observar que, com os dispositivos de transparência disponíveis desde a criação da Lei de Acesso à Informação (LAI), em 2012, a população passou a acompanhar mais os dados governamentais e surgiram projetos de análise destas informações disponibilizadas, como o projeto *Operação Serenata de Amor*, idealizado por [Musskopf, Cuducos e Cabral \(2016\)](#). Porém, o projeto só foi possível com a disponibilização das informações num formato legível por máquina, então é válido o questionamento da efetividade dos dispositivos de governo eletrônico tradicionais para melhoria do processo de transparência das instituições, pois, embora o acesso tenha aumentado, grande parte da população continua sem acompanhar as ações governamentais.

Paralelo ao surgimento do governo eletrônico e os dispositivos de transparência, o desenvolvimento das tecnologias de informação e comunicação tornou possível a captação de informações geradas nas cidades, e com o uso de técnicas cada vez mais efetivas de análise de dados pode-se consumir os recursos públicos com mais eficiência e eficácia.

Esta combinação de cenários evidencia um paradoxo, entre o desenvolvimento tecnológico disponível, e o lento desenvolvimento político-social, mas este pode apresentar melhora se as ferramentas de governo eletrônico disponibilizadas favorecerem o uso de tecnologias de informação inteligentes, e análise de dados. Para preencher as lacunas deixadas pelos portais de transparência atuais, e incentivar a análise científica dos dados institucionais públicos, surgem os portais de dados abertos que têm a finalidade de padronizar e sistematizar a publicação de informações públicas de forma que elas possam ser lidas tanto por humanos, quanto por

máquinas.

Entretanto, o processo de abertura das informações públicas de uma instituição não é tarefa trivial, e deve atender uma série de diretrizes legais e institucionais para que a divulgação esteja livre de questionamento quanto à consistência e lisura.

1.1 Justificativa

A publicação de dados de instituições públicas é um tema que vem sendo aprimorado nas leis desde quatro de maio do ano 2000, com o artigo 48 da Lei complementar 101 que determina a adoção de instrumentos de transparência na gestão fiscal em meios eletrônicos de acesso público para as informações orçamentárias e de prestação de contas. Apesar disso, ainda encontramos entraves para abertura dos dados institucionais nas instituições governamentais em geral, e a situação não é diferente nas Instituições de Ensino Superior *IES*. A *UFS* possui dados e informações semelhantes, não somente às *IES*, mas a qualquer outra instituição pública federal, principalmente no âmbito financeiro e contábil, dessa forma, o processo de abertura dos dados para estas instituições deve ser bastante similar. Mesmo assim, a *UFS* iniciou o processo com atraso em relação às outras *IES*. não por desconhecimento da lei, mas por não conseguir desmistificar o processo de abertura em tempo hábil, em outras palavras, por não existir um procedimento sistemático devidamente documentado, que facilitasse a execução da tarefa.

Além da motivação legal, é evidente a demanda por justiça e combate à corrupção na administração pública do país. Para atender a este clamor social, uma das abordagens possíveis é fazer uso de técnicas de ciência de dados com as informações disponibilizadas pelo governo. Entretanto, os atuais portais de transparência são falhos em disponibilizar informações que possam ser lidas automaticamente por aplicativos e sistemas de informação, pois:

- Falta padronização para disponibilizar as informações dentre os portais de transparência do Brasil.
- Não existe a opção de *download* em um formato estruturado para ser lido por máquina. As informações, na maioria das vezes, são disponibilizadas em formato não-estruturado (documentos *PDF* ou páginas *HTML*, por exemplo), o que dificulta a extração da informação e leitura por máquina para análise de dados.

De acordo com [Attard et al. \(2015\)](#), as iniciativas de governo aberto para fortalecimento da democracia são baseadas em três pilares: transparência, participação cidadã e colaboração. Portanto, as duas características dos portais de transparência, anteriormente listadas, mostram que eles são limitados no fortalecimento da democracia, pois não atendem com eficiência os pilares da participação cidadã e da colaboração. Além disso, estas duas características apresentadas anteriormente dificultam o processo de extração automática das informações nos portais de

transparência, e consequentemente da análise padronizada das informações. Somada à falta de padronização, os portais de transparência não possuem uma linguagem amigável e convidativa à população, o que dificulta a sua popularização entre as pessoas que não lidam com o linguajar administrativo, financeiro, ou contábil, e consequentemente as afastando da sua obrigação civil de acompanhar as ações governamentais.

Durante a pesquisa realizada, não foram encontradas soluções adequadas que satisfizessem essas lacunas necessárias ao cumprimento legal e social das *IES*. Portanto, diante do cenário configurado, justifica-se a realização deste trabalho que tem como objetivo propôr uma arquitetura de dados abertos que auxilie no processo de abertura dos dados institucionais, através do desenvolvimento de um *Plano de Dados Abertos (PDA)*, e aplicação deste com a finalidade de disponibilizar os dados institucionais no Portal de Dados Abertos.

1.2 Objetivo Geral

O objetivo geral deste trabalho consiste em propôr uma arquitetura de dados abertos, em conformidade com a norma *ISO/IEC/IEEE 42010:2011*, para a Universidade Federal de Sergipe, que sirva de base para o seu processo de abertura de dados institucionais públicos, contribuindo com o processo de transparência da instituição.

1.3 Objetivos Específicos

Para atender o objetivo geral, foram estabelecidos os seguintes objetivos específicos:

- Análise dos conceitos de dados abertos, com o objetivo de afirmar a importância de se desenvolver uma política de dados abertos em instituições públicas de ensino federais.
- Identificação dos desafios de desenvolver uma política de dados abertos para instituições públicas.
- Aplicando a arquitetura desenvolvida, elaborar um Plano de Dados Abertos (*PDA*) para orientar a abertura de dados da Universidade Federal de Sergipe, de forma a institucionalizar uma política de publicação de dados públicos.
- Desenvolvimento de um Portal de Dados Abertos aplicando o *PDA* elaborado.
- Desenvolver bibliotecas reutilizáveis que possam extrair e publicar *datasets* no portal de dados abertos, por meio de interfaces *RESTful*.

1.4 Hipóteses

Hipótese 1: O desenvolvimento de uma arquitetura de dados abertos provê a identificação de partes interessadas e seus respectivos interesses, requisitos do sistema, e diferentes pontos de vista do negócio. O conhecimento destes atributos torna possível a elaboração do plano de dados abertos.

Hipótese 2: Através de um plano de abertura de dados é possível orientar e facilitar o processo de abertura de dados em instituições públicas de ensino superior.

1.5 Metodologia de Pesquisa

De acordo com [Silva e Menezes \(2001\)](#), a classificação da pesquisa para este trabalho pode ser descrita da seguinte forma:

- **Quanto à natureza:** Aplicada, pois tem o objetivo de gerar conhecimento para aplicação prática de uma solução, com o objetivo de resolver um problema. Este trabalho propôs uma arquitetura de dados abertos para a instituição, com o intuito de dar mais transparência para os dados da instituição e atender dispositivos exigidos pela legislação.
- **Quanto a abordagem do problema:** Qualitativa, pois o ambiente natural é a fonte direta para coleta de dados. O processo e seu significado são os focos principais de abordagem.
- **Quanto aos objetivos:** Exploratória, pois visa aumentar a familiaridade com o problema, com o objetivo de torná-lo explícito, ou a construir hipóteses. Este trabalho buscou explorar as bases de dados, assim como seus *datasets* a fim de estruturá-los e disponibilizá-los para a população.
- **Quanto aos procedimentos técnicos:** É um estudo de caso, pois envolve o estudo profundo do problema, com o objetivo de permitir seu amplo conhecimento. Durante a execução do trabalho técnico de extração das informações, algumas técnicas de seleção e tratamento dos dados foram aplicados:
 - Inclusão dos *datasets* homologados pelos usuários.
 - Suspensão dos *datasets* que não puderam ser homologados por um analista de sistemas da área.
 - Suspensão dos *datasets* não alimentados pelos usuários.
 - Tratamento dos dados para evitar a inclusão de dados sensíveis dos cidadãos e da instituição.

1.6 Trabalhos Relacionados

Inicialmente, Uma revisão da literatura foi realizada com o propósito de encontrar trabalhos relacionados com os temas: dados abertos, transparência governamental, e arquitetura de dados abertos. Foram utilizadas as bases:

- **Scopus**¹
- **ACM Digital Library**²
- **IEEE Xplorer Digital Library**³
- **Springer**⁴
- **Google Scholar**⁵

Para realizar a pesquisa, foram selecionados os trabalhos publicados entre o período de 2008 a 2018, com o objetivo de manter a pesquisa com um período de tempo contemporâneo, e utilizados os termos: *open data*, *open data portal*, *open data architecture*, *open government data*, e *transparency*.

Como resultado, foram obtidos muitos trabalhos sobre os temas, mas nenhum relacionado especificamente sobre o desenvolvimento de uma arquitetura de dados abertos governamental, com foco no desenvolvimento de um plano de dados abertos. Entretanto, dois manuais de publicação de dados foram encontrados na Internet: o *Manual para Elaboração de Plano de Dados Abertos* desenvolvido pelo [MPOG \(2015\)](#); e o *Open Data Field Guide* desenvolvido pela [Socrata \(2015\)](#).

Com relação à arquitetura, alguns trabalhos foram encontrados, porém, alguns descreviam apenas do uso de ferramentas *ETL* - (*Extract, Transform, Load*) para o processo de extração dos dados, sem citar maiores detalhes. Outros apresentavam abordagens que envolviam atuação humana no processo. Posteriormente, visitando o *Google Scholar* em busca, aleatoriamente, de outros trabalhos sobre dados abertos, foi encontrada a dissertação de mestrado de [Kong \(2015\)](#).

Por fim, sobre portais de dados abertos, foi encontrada uma revisão sistemática da literatura sobre as iniciativas de portais de dados abertos pelo mundo.

Os trabalhos relacionados ao tema são:

- [Possamai \(2016\)](#) em sua tese de doutorado dedica-se ao estudo dos fatores críticos para adoção de uma iniciativa de dados abertos governamentais como referência de tratamento

¹ *Scopus* - <<https://www.scopus.com>>

² *ACM Digital Library* - <<http://dl.acm.org/>>

³ *IEEE Xplorer Digital Library* - <<http://ieeexplore.ieee.org/Xplore/home.jsp>>

⁴ *Springer* - <<https://link.springer.com/>>

⁵ *Google Scholar* - <<https://scholar.google.com>>

das informações públicas na Era Digital. Em sua pesquisa emprega uma metodologia qualitativa e o método de estudo exploratório da política de dados abertos do Governo Federal, que foi iniciada e implementada pela INDA⁶. Sua pesquisa contribui com subsídios teóricos-práticos para o processo de aprendizagem e planejamento de estratégias de publicação de dados abertos adaptadas aos recursos disponíveis. Este trabalho teve grande relevância nesta dissertação pois ressalta a importância de estabelecer padrões abertos de interoperabilidade na administração pública, além de reafirmar a relevância dos dados abertos como uma política de governo eletrônico.

- [Attard et al. \(2015\)](#) realizaram uma revisão sistemática da literatura com o objetivo de catalogar as iniciativas de dados abertos governamentais, além de discutirem sobre o ciclo de vida dos dados abertos, sobre o processo de abertura dos dados dentro das organizações governamentais, sobre os principais problemas que impedem o processo de abertura dos dados, e avaliarem as iniciativas de dados abertos catalogadas em seus estudos primários. Tem relevância para este trabalho, pois mostra que as iniciativas de dados abertos pelo mundo estão crescendo como alternativa viável de governo eletrônico.
- [Kong \(2015\)](#) desenvolveu uma arquitetura para publicação de dados abertos, porém com uma abordagem diferente deste trabalho, onde uma ferramenta para extração e publicação das informações foi desenvolvida, pois a fonte das informações é uma *API RESTful*. [Kong \(2015\)](#) utiliza uma ferramenta *ETL* para extração dos dados, oriundos de uma base hierárquica, do *Sistema Integrado de Administração de Serviços Gerais (SIASG)*, e carga dos dados numa base relacional de avaliação de qualidade destes.
- [Correa et al. \(2018\)](#) desenvolveu uma arquitetura, baseada no método *ADD (Attribute Driven Design)*, para publicação de dados abertos através da extração de dados oriundos dos portais de transparência convencionais. Relaciona-se com o presente trabalho pois extrai as informações dos portais de transparência (soluções de governo eletrônico já existentes), porém, devido a falta de padronização dos portais de transparência, cada portal será um caso de uso diferente, o que torna a solução pouco escalável.

1.7 Organização da Dissertação

Os capítulos que se seguem desta dissertação estão dispostos conforme a descrição a seguir:

- O capítulo 2 apresenta a fundamentação teórica que subsidia este trabalho. São apresentados os conceitos de transparência governamental, cidades inteligentes e Internet das Coisas, dados abertos, a ferramenta *CKAN*, a *ISO/IEC/IEEE 42010:2011*, diagramas de componentes e arquiteturas *REST*.

⁶ Infraestrutura Nacional de Dados Abertos

- O capítulo 3 descreve a arquitetura *Open Data UFS* identificando as partes interessadas e seus respectivos interesses, visões, requisitos funcionais e não funcionais, modelos que podem ser utilizados em cada visão, regras de correspondência entre visões, diagrama de componentes *UML* representando a visão da aplicação, procedimentos de personalização e instalação do *CKAN* no servidor e tecnologias envolvidas, descrição da etapa de desenvolvimento da solução de extração e carregamento dos dados no portal, recursos desenvolvidos para a *API Sistemas*, e por fim as dificuldades encontradas durante o desenvolvimento da arquitetura.
- O capítulo 4 apresenta o Plano de Dados Abertos desenvolvido para guiar o processo de abertura dos dados da instituição, e o Portal de Dados Abertos, e suas funcionalidades, desenvolvido durante o projeto.
- O capítulo 5 por fim, conclui o trabalho desenvolvido, retomando as principais contribuições, dispondo as limitações do projeto, e sugerindo trabalhos futuros.

2

Fundamentação Teórica

Neste capítulo serão apresentados os conceitos teóricos que compõem o trabalho. Na Seção 2.1, o conceito de transparência é definido, e a eficiência do modelo de transparência tradicional, adotado até então, é questionada. Em seguida, são apresentados os conceitos de Cidades Inteligentes e Internet das Coisas (*IoT*), na Seção 2.2. Na Seção 2.3 o conceito de Dados Abertos é definido, consolidando o tema principal do trabalho, e na Seção 2.4 o software de publicação de dados *CKAN*, utilizado no trabalho, é apresentado ao leitor. Na Seção 2.5 a norma *ISO/IEC/IEEE 42010:2011* é apresentada com o intuito de esclarecer os conceitos relativos à descrição de arquiteturas de *software* corporativas. Com o objetivo de esclarecer a simbologia utilizada no capítulo 3, uma breve introdução sobre Diagrama de Componentes foi elaborada na Seção 2.6. Por fim, para concluir o entendimento sobre *API's RESTful*, citadas no trabalho, a Seção 2.7 apresenta os conceitos do padrão arquitetural *REST*.

2.1 Transparência Governamental

De acordo com Possamai (2016), em meados da primeira década do século XXI, com o surgimento da web 2.0, os governos e agentes públicos ensaiaram uma ampliação no contato com o cidadão por meio das mídias sociais e redes de compartilhamento. Dessa forma, surgiu o "Governo 2.0" com os cidadãos cada vez mais participativos na Web. Em consequência da mudança na forma como interagiam na rede, ampliou-se o debate acerca da promoção de um governo transparente e com os seus cidadãos participativos. A partir desse momento, o governo aberto e a transparência públicas passaram a ser pauta da agenda de políticas públicas.

As organizações moldam as suas políticas de transparência como prática de governança corporativa em resposta a pressões exercidas por organizações que detêm poder sobre outras, por meio de modelos amplamente aceitos com o objetivo de normatizar condições e métodos de trabalho (por trabalho, entenda-se a atividade monitorada pela comunidade). Mas apesar

de parecer sólido e consensual, o conceito de transparência pode ter significado diferente a depender de quem estabelece a sua legitimidade. Os EUA e o Reino Unido, por exemplo, são países referenciais no assunto, mas ainda possuem muita informação contida apesar da posição de liderança. (BARBALHO, 2014)

Segundo Dawes e Helbig (2010), as iniciativas de transparência, normalmente, possuem como objetivo fornecer aos cidadãos e outras partes interessadas um meio de vigiar e fiscalizar as ações governamentais e agentes públicos. Mas além disso, dados governamentais podem ser publicados à população com o intuito de gerar valor agregado (social e econômico) a quem os consome. Neste último objetivo, o autor refere-se aos portais de transparência, mas o conceito é perfeitamente aplicável aos portais de publicação de dados abertos, cujo conceito será introduzido adiante.

Porém, conforme observado por Murillo (2015), na América Latina, a corrupção e a ineficiência, normalmente presentes nas instituições políticas, ainda são um desafio para as futuras gerações. Mesmo com a utilização de dispositivos legais de transparência e de tecnologias de informação e comunicação modernas (portais de transparência) que incentivam a vigilância do patrimônio público pela população, no Brasil, estes mecanismos não favorecem dois dos três pilares de fortalecimento da democracia (ATTARD et al., 2015): participação cidadã e colaboração. O Brasil, que é considerado o país com maior índice de abertura de dados governamentais na América Latina, ainda possui alto índice de percepção da corrupção de acordo com a Transparency International (2018).

2.2 Cidades Inteligentes e Internet das Coisas (IoT)

Não há uma definição única para o termo "cidades inteligentes", e por vezes ele pode ser referido substituindo-se o termo "inteligente" por "digital", ou a palavra cidades por "ambiente".

O trabalho de Albino, Berardi e Dangelico (2015) compila várias definições importantes publicadas para cidades inteligentes, dentre elas está a de Bakıcı, Almirall e Wareham (2013) que define o termo como um ambiente tecnológico que conecta pessoas informações e elementos da cidade utilizando novas tecnologias em favor de um ambiente sustentável, comercialmente competitivo e inovador, que agrega qualidade de vida aos cidadãos. A definição de Chen (2010) completa a definição anterior, ao passo que cita redes de sensores sem fio como uma importante tecnologia na infraestrutura das cidades otimizando operações logísticas que dão suporte à vida diária aumentando a qualidade de vida dos cidadãos. A definição de cidades inteligentes descrita por Kourtit e Nijkamp (2012) aborda o tema com uma visão patrimonial, descrevendo as cidades inteligentes como o resultado da junção de estratégias criativas e voltadas ao conhecimento melhorando os desempenhos competitivos, socio-econômicas, ecológicas e logísticas. São, em termos práticos, o resultado dos nossos patrimônios ou capitais humano, infraestrutural, social e empreendedor.

Em síntese, podemos definir uma cidade inteligente como um ambiente de convívio de pessoas com uma delimitação geográfica e uma gestão em comum, e que utiliza tecnologia agregada aos seus elementos para conectá-los e captar dados que podem ser utilizados posteriormente para gerar capital infraestrutural, valorizando o capital social e humano para melhoria dos serviços que dão suporte aos cidadãos no dia a dia.

Internet das coisas ou *IoT*, pode ser compreendido como um paradigma moderno da telecomunicação sem fio, cuja idéia básica é a presença generalizada de dispositivos com tags de radiofrequência (RFID), sensores, atuadores, e *smartphones* interagindo e cooperando entre si com a finalidade de atingir objetivos em comum. Este conceito traz um alto impacto na vida cotidiana dos cidadãos, ao passo que estará presente em várias áreas domésticas e laborais como vida assistida, automação residencial e aprendizado aprimorado. (ATZORI; IERA; MORABITO, 2010)

Dentro de cidades inteligentes, a utilização de dispositivos *IoT* é a combinação ideal entre a necessidade de resolução de problemas dos ambientes urbanos e o uso eficiente de *TIC*. Os cenários possíveis para utilização são os mais variados: monitoramento de tráfego automotivo através de sensores espalhados pelas vias; controle de consumo residencial de energia elétrica; e troca de informações entre eletrodomésticos para otimização de tarefas rotineiras são alguns exemplos citados por Tomas (2014).

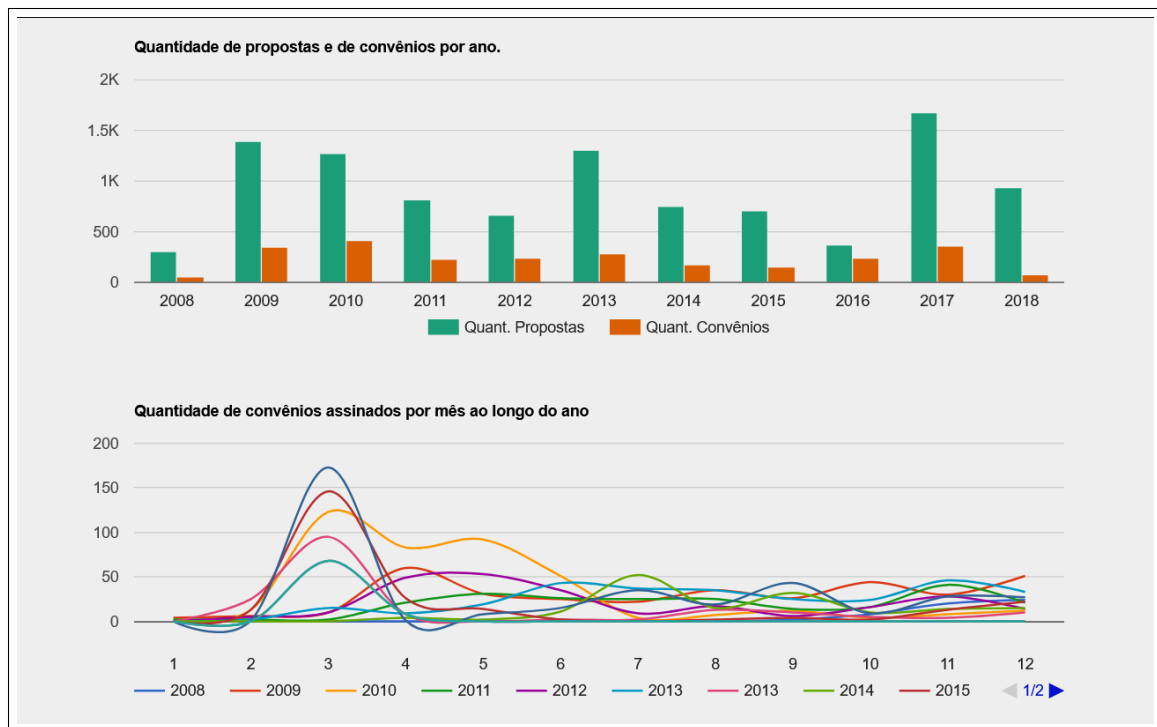
De acordo com Atzori, Iera e Morabito (2010), a Internet revolucionou a maneira como vivemos promovendo interações virtuais, nos âmbitos social e profissional, entre as pessoas. A *IoT* agrega uma nova dimensão a este processo, permitindo as comunicações com e entre objetos inteligentes, tornando a Internet e as suas interações mais presentes e integradas ao dia a dia e, para isso, *IoT* deve ser considerada como parte da Internet global do futuro.

2.3 Dados Abertos

A política de publicação de dados abertos terá um papel de extrema relevância para que o Brasil cumpra os três pilares de fortalecimento da democracia (ATTARD et al., 2015). Além disso, os portais de dados abertos serão os disseminadores dos dados captados pelos dispositivos *IoT*, em cidades inteligentes, para serem utilizados por cidadãos engajados em melhorar a qualidade de vida nas cidades, e otimizar os seus recursos.

Como pode-se observar na Seção 2.1 sobre Transparência Governamental, apenas a divulgação das informações não é suficiente para garantir a transparência. O Brasil, como previamente afirmado, ainda não tira total proveito das tecnologias disponíveis, pois apesar de disponibilizar as informações, muitas delas estão registradas sob conceitos ou linguagem que a média da população não domina. São, portanto, necessárias ferramentas que traduzam a linguagem administrativa, financeira, contábil, jurídica e burocrática aos cidadãos em geral, a exemplo do portal *Transparência Traduzida* (CAMPOS et al., 2018), ilustrado na Figura 1.

Figura 1 – Portal Transparência Traduzida



Fonte - (CAMPOS et al., 2018)

Mas como fazê-lo, e automatizá-lo, utilizando as tecnologias disponíveis, quando as informações não estão nem ao menos padronizadas, ou sequer em formatos que possam ser lidas por máquina? Diante deste contexto, os portais de dados abertos suprem essa lacuna de transparência, pois visam publicação de dados em formatos de preferência abertos, padronizados, se possível contextualizados, e que podem ser lidos por máquina.

O conceito de dados abertos baseia-se no fato de que existem certas informações cujo acesso deve estar acima de copyright, patentes, censura ou qualquer outro acesso privado (BERTOT et al., 2014). De acordo com o Open Data Handbook (OKI, 2011), mantido pelo *Open Knowledge International*, dados abertos podem ser definidos como informações que podem ser usadas, reutilizadas e redistribuídas sem restrição, por qualquer pessoa, e estão sujeitos a no máximo duas regras: atribuição da fonte ao autor, e o compartilhamento das informações está sujeito às mesmas regras da publicação.

De acordo com a OKI (2011), é extremamente importante definir três características fundamentais do que é ser aberto, com o objetivo de valorizar a interoperabilidade e garantir que a parte livre da informação possa ser cruzada com outras informações igualmente livres.

- **Disponibilidade e Acesso:** Os dados devem ser disponibilizados como um todo, num formato conveniente e modificável, preferencialmente online e sob custo não maior que o custo razoável da reprodução.
- **Reutilização e Redistribuição:** Os dados devem ser disponibilizados sob termos claros

que permitam a sua reutilização, redistribuição e cruzamento com outros dados.

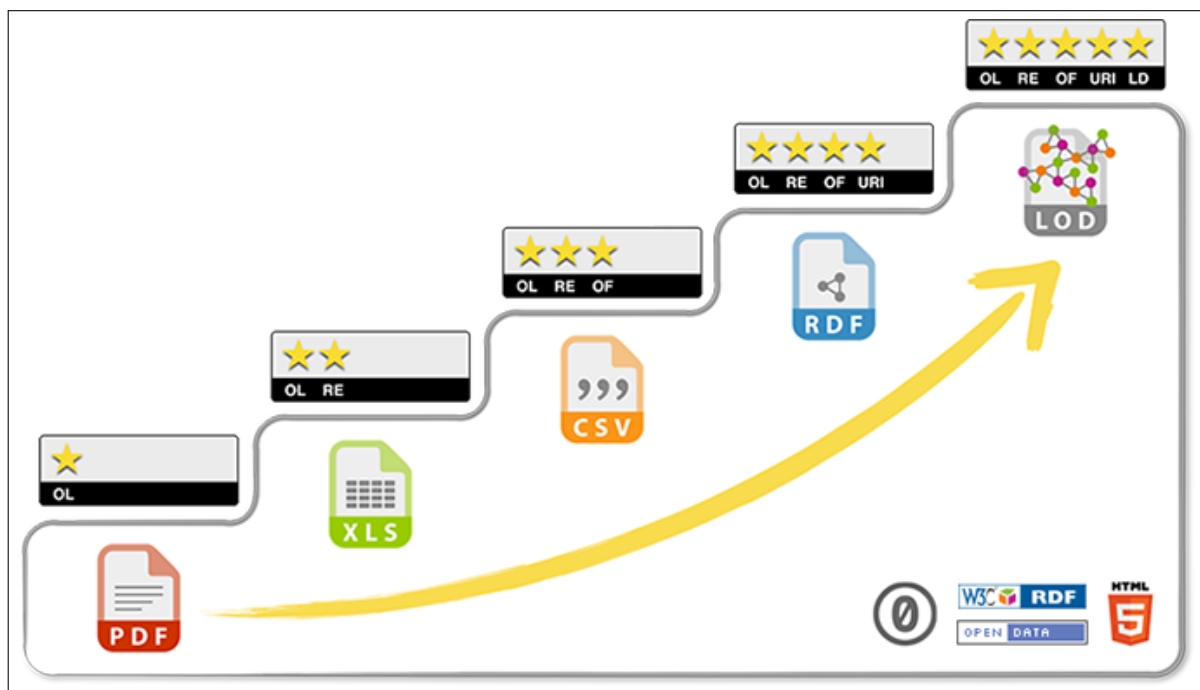
- **Participação Universal:** Qualquer informação deve ser passível de utilização, reutilização e redistribuição, sem qualquer discriminação de área de atuação, pessoa ou grupo.

De acordo com o Portal Brasileiro de Dados Abertos¹ do Governo Federal, em 2009, o ativista e especialista em políticas públicas, David Eaves, propôs três leis que visam garantir a abertura e indexação dos dados abertos. São elas:

1. Se o dado não puder ser encontrado e indexado na web, ele não existe.
2. Se o dado não estiver em formato aberto e compreensível por máquina não pode ser reaproveitado.
3. Um dispositivo legal deve garantir a replicação do dado, com o objetivo de torná-lo disponível e útil à comunidade.

De forma a encorajar as pessoas, principalmente os donos dos dados governamentais, a melhorar a qualidade dos dados disponibilizados, Tim Berners-Lee, considerado o criador da WWW (*World Wide Web*), definiu o modelo *5 estrelas* de classificação das publicações de dados abertos online.

Figura 2 – Escada de evolução no modelo 5 estrelas de publicação de dados abertos, definido por Tim Berners-Lee



Fonte - (HAUSENBLAS; KIM, 2012)

¹ Portal Brasileiro de Dados Abertos - <<http://dados.gov.br/>>

- **1 estrela:** Disponível na web em qualquer formato, mas com uma licença aberta para que seja considerado aberto;
- **2 estrelas:** Disponível numa estrutura de dados que pode ser lida por máquina (uma planilha no formato *XLS*, por exemplo, ao invés de uma imagem escaneada de uma folha de papel com a informação);
- **3 estrelas:** O mesmo que o item 2 estrelas, porém num formato não-proprietário (*CSV*, ao invés de *XLS*);
- **4 estrelas:** Deve satisfazer todas as condições anteriores e, além disso, deve utilizar padrões abertos do *W3C* (*RDF* e *SPARQL*) para identificar os dados, dessa forma, outras pessoas podem referenciá-los;
- **5 estrelas:** Deve satisfazer todas as condições anteriores e, além disso, ter os dados conectados com os dados de outros portais ou repositórios;

2.4 CKAN

De acordo com [Owiecińska e Legierski \(2015\)](#), o *CKAN* (*Comprehensive Knowledge Archive Network*) é um portal de publicação de dados, de código aberto, e distribuído sob a licença *open source* Affero GNU GPL v3.0. Mantido pela *The Open Knowledge Foundation* desde o início do projeto, em março de 2006, tornou-se a plataforma líder de dados abertos, com várias instâncias instaladas em todo o mundo. Pode-se citar, por exemplo, os portais:

- [data.gov](#) - Portal de dados abertos do governo norte-americano
- [data.gov.uk](#) - Portal de dados abertos do governo britânico
- [dados.gov.br](#) - Portal de dados abertos do governo brasileiro
- [dados.al.gov.br](#) - Portal de dados abertos do Governo de Alagoas
- [web.transparencia.pe.gov.br/dados-abertos](#) - Portal de dados abertos do governo de Pernambuco. Uma curiosidade sobre este portal é que ele é referenciado na página principal do portal de transparência do governo de Pernambuco, agregando o conceito de dados abertos ao da transparência tradicional.
- [dados.recife.pe.gov.br](#) - Portal de dados abertos da prefeitura de Recife.
- [dados.ufrn.br](#) - Portal de dados abertos da UFRN (Universidade Federal do Rio Grande do Norte).
- [opendata.smart.ufs.br](#) - Portal de dados abertos da UFS desenvolvido academicamente durante a confecção deste trabalho.

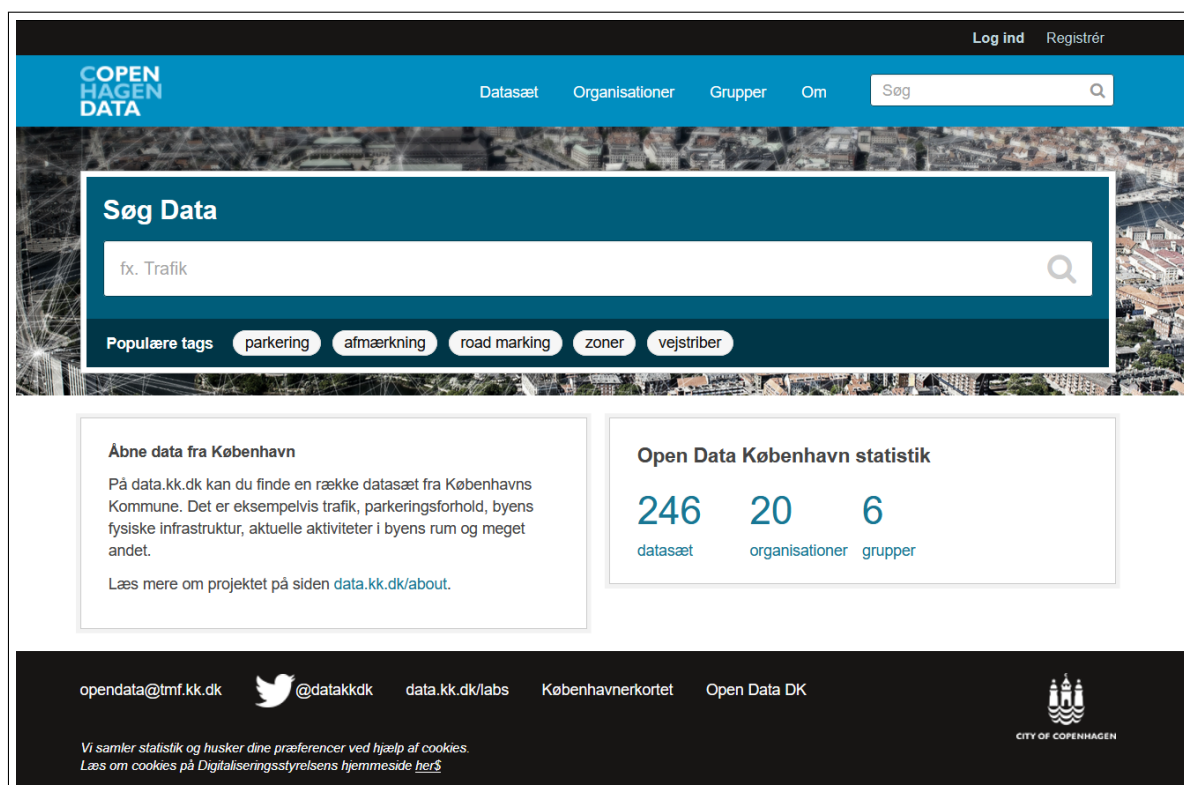
Outras instâncias do CKAN podem ser consultadas no seu site².

Figura 3 – Portal Alagoas em Dados e Informações



Fonte - SEPLAG-AL (2013)

Figura 4 – Portal Copenhagen Data



Fonte - Copenhagen Technology and Environment Administration (2013)

² CKAN Instances - <<https://ckan.org/about/instances/>>

O CKAN, por ser um software livre, permite que seu código seja personalizado de acordo com as necessidades do utilizador. Dessa forma, pode-se encontrar portais de dados abertos pelo mundo, personalizados a partir dele, com interfaces de apresentação bem diferentes, como pode-se observar nas Figuras 3 e 4.

O CKAN permite armazenar e gerenciar repositório de dados abertos, como também a publicação, pesquisa e visualização de conjuntos de dados. Além disso, também permite a pesquisa de datasets, *download* e *upload* de conjuntos de dados em vários formatos (*JSON*, *CSV*, *XLS*, *XLSX*, *ODS*, *PDF*, dentre outros) através de uma *API REST* disponibilizada à comunidade.

O CKAN é utilizado por governos locais e nacionais, instituições de pesquisa e outras organizações coletoras de dados. É um software de código aberto, desenvolvido em *Python*, e com uma comunidade ativa de colaboradores que mantêm sua estrutura central, mas possui uma comunidade ainda maior criando bibliotecas de extensão para ele ([Open Knowledge International, 2006](#)).

Nem todos os portais de dados abertos são desenvolvidos utilizando o CKAN. Alguns portais são desenvolvidos a partir de outros *softwares*, ou como simples repositórios de dados, como os exemplos a seguir:

- [brasil.io](#) - Repositório independente de dados abertos de vários setores do país, mantido por Álvaro Justen e colaboradores.
- [kaggle.com](#) - É descrito como uma comunidade online de cientistas de dados e pesquisadores de aprendizado de máquina, do *Google*. A comunidade é conhecida também por promover competições de ciência de dados utilizando seus repositórios.

Para desenvolver o Portal *Dados Abertos UFS*, o CKAN foi escolhido pois além de ser muito utilizado no mundo inteiro, e possuir uma grande comunidade ativa de administradores de dados mantendo instancias de seus portais baseadas nele, este também é recomendado pela plataforma de cidades inteligentes *FIWARE*³, que é uma infraestrutura em nuvem de plataforma aberta para *IoT*, financiada pela Comissão Européia e pela União Européia, baseada no *OpenStack*⁴ (plataforma *open source* para computação em nuvem). Atualmente, no *DCOMP*, existe um projeto para implantação do *FIWARE* em andamento, e vislumbrando a inserção da *UFS* no contexto de cidades inteligentes, e a consequente publicação de dados de seus sensores *IoT*, o CKAN era a escolha natural para o desenvolvimento deste trabalho.

³ [<https://www.fiware.org/>](https://www.fiware.org/)

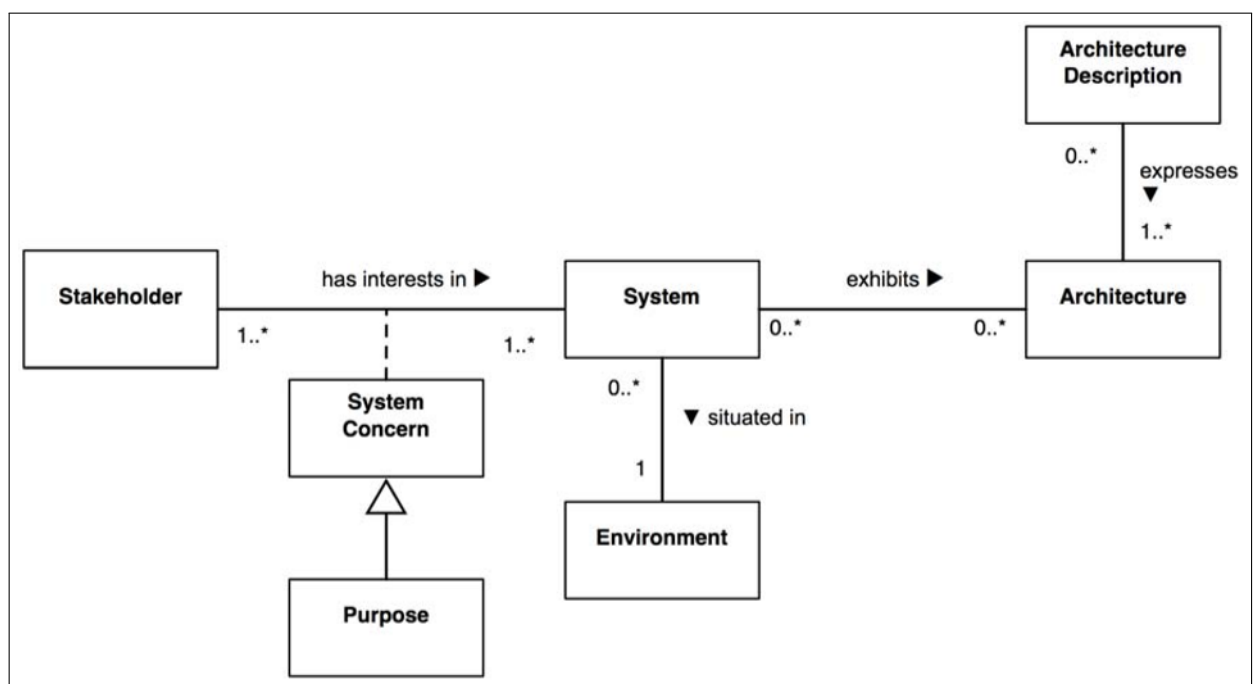
⁴ [<https://www.openstack.org/>](https://www.openstack.org/)

2.5 Norma ISO/IEC/IEEE 42010:2011

De acordo com a [International Organization Of Standardization \(2011\)](#), a norma ISO/IEC/IEEE 42010:2011 diz que a descrição de arquitetura gera uma ou mais arquiteturas de sistemas e softwares, que devem conter as partes interessadas (*stakeholders*) e seus interesses (*concerns*). Dessa forma, os sistemas e softwares podem alcançar o seu devido propósito no ambiente.

2.5.1 Contexto e Modelo Conceitual da Descrição de uma Arquitetura

Figura 5 – Contexto da Descrição de uma Arquitetura



Fonte - ([International Organization Of Standardization, 2011](#))

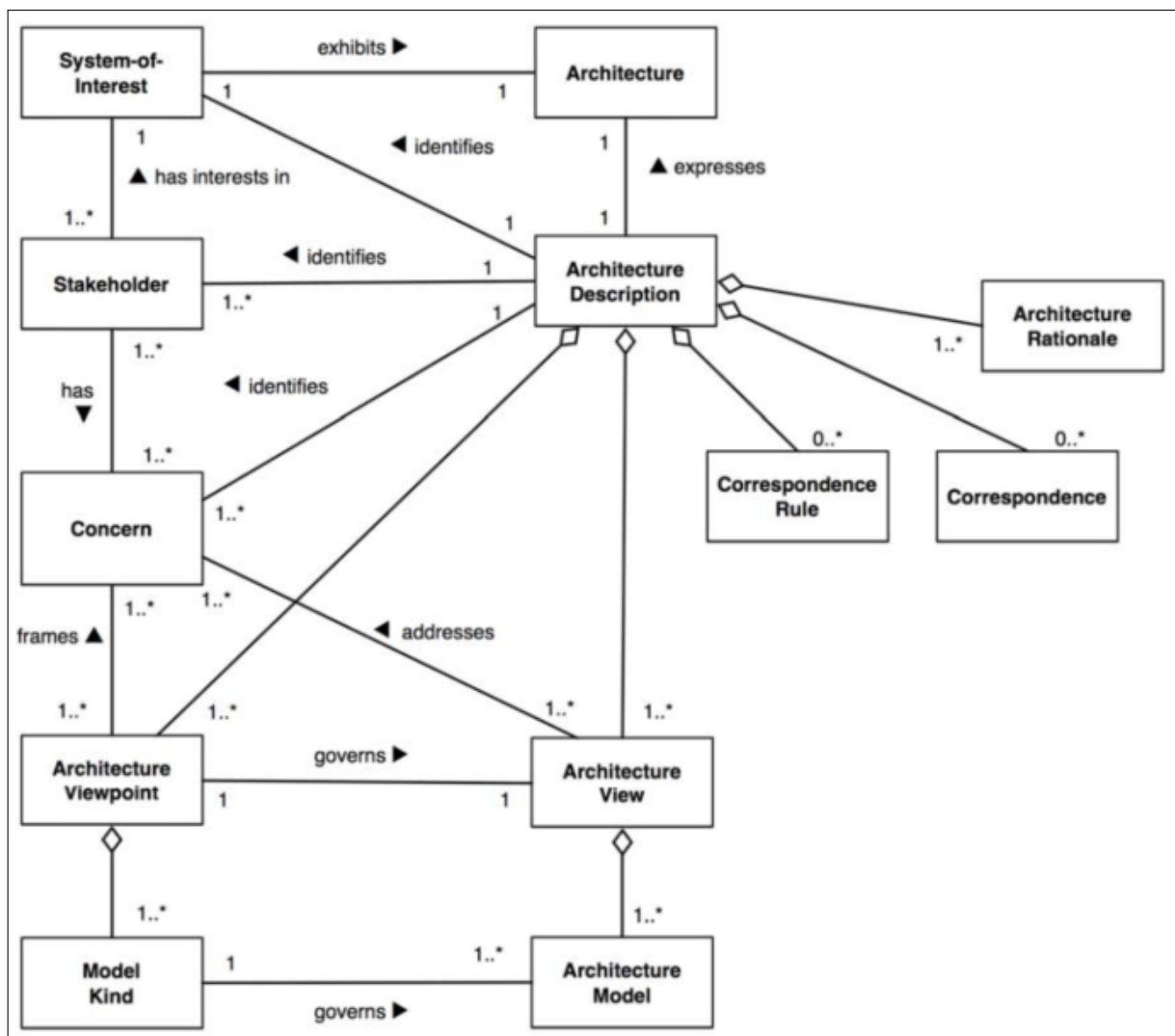
É importante ressaltar a diferença entre sistema e software, pois de acordo com a norma, o sistema é uma obra humana composta por um ou mais dos itens a seguir, dentre as quais o software se inclui: hardware, software, dados, pessoas, processos, procedimentos, instalações, materiais, e entidades naturais.

Segundo a norma, o processo de arquitetar (conceber a arquitetura) ocorre durante todo o ciclo de vida do sistema, e não somente em uma determinada fase do ciclo. Portanto, a descrição de uma arquitetura é produto de trabalho dos sistemas e arquiteturas de software e, consequentemente, um produto resultante de toda a execução das atividades que compõem um sistema. Dessa forma, a descrição de uma arquitetura:

- Deve identificar sistemas e pessoas, e seus respectivos interesses.

- Deve definir uma visão arquitetural a partir de um ou mais modelos arquiteturais.
- Possui elementos, que são cada pessoa interessada (*stakeholder*), como também seus interesses, pontos de vista, visões arquiteturais, modelos arquiteturais, decisões arquiteturais, e lógica envolvida.
- Documenta as decisões arquiteturais e lógica envolvida.

Figura 6 – Modelo Conceitual da Descrição de uma Arquitetura



Fonte - ([International Organization Of Standardization, 2011](#))

A descrição de uma arquitetura de sistemas possui as várias finalidades a seguir:

- Servir de base para o projeto do sistema e suas atividades de desenvolvimento.
- Servir para analisar e avaliar implementações alternativas de uma arquitetura.
- Servir de documentação de desenvolvimento e manutenção.

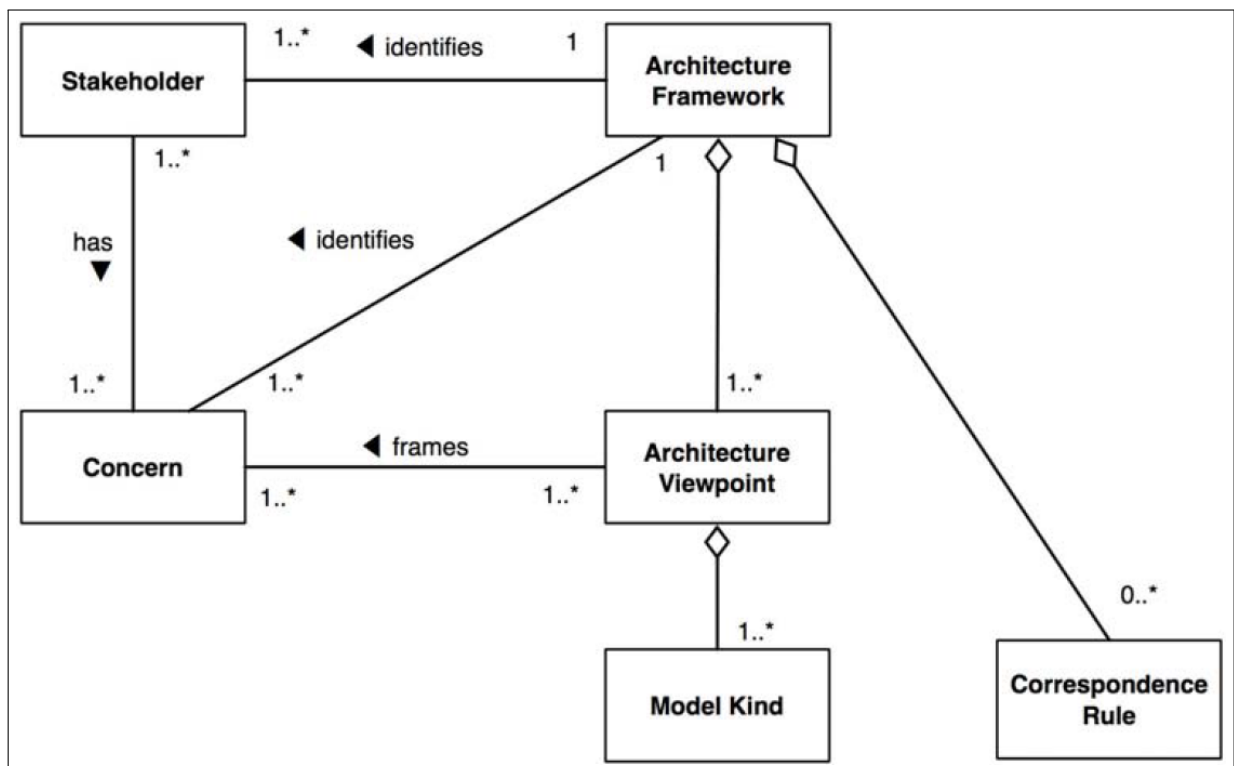
- Documentar aspectos essenciais do sistema como pretensão de uso, princípios limitações, pontos de flexibilidade, aspectos de mudanças futuras, decisões arquiteturas e suas implicações lógicas.
- Servir como ponto de entrada para ferramentas de automação para simulação, geração e análise de sistemas.
- Especificar grupos de sistemas com funcionalidades em comum.
- Servir de comunicação entre os grupos envolvidos no desenvolvimento, produção, implantação, operação, e manutenção do sistema.
- Servir de base para confecção de documentos correlatos ao sistema.
- Servir de comunicação entre clientes, compradores, fornecedores e desenvolvedores como parte das negociações contratuais.
- Servir de documentação das características, funcionalidades e projeto para potenciais clientes, compradores, proprietários, operadores e agregadores.
- Servir de plano para a transição de uma arquitetura legada para uma nova arquitetura.
- Servir de manual para suporte operacional e de infraestrutura, e gerenciamento de configuração.
- Auxiliar nas atividade de planejamento, programação e despesas do sistema.
- Estabelecer critérios para a implementações certificadas em conformidade com a arquitetura.
- Servir de mecanismo para para políticas externas, de projeto, e/ou internas da organização.
- Servir de base para revisão, análise e avaliação do sistema durante o seu ciclo de vida.
- Servir de base para avaliar arquiteturas alternativas.
- Compartilhamento de lições aprendidas e reuso do conhecimento arquitetural através de pontos de opiniões, padrões e estilos.
- treinamento e aprendizado em melhores práticas ao planejar a arquitetura e evolução do sistema, para os interessados (*stakeholders*), e outros grupos.

2.5.2 Frameworks e Linguagem de Descrição de uma Arquitetura

Segundo a norma ISO/IEC/IEEE 42010:2011, da [International Organization Of Standardization \(2011\)](#), frameworks e linguagens de descrição de uma arquitetura são construídos alicerçados nos conceitos da descrição da arquitetura presentes na norma.

O *framework* de uma arquitetura estabelece uma prática comum para criar, interpretar, analisar e utilizar as descrições da arquitetura dentro e um domínio particular ou comunidade de interessados, de acordo com a norma. A Figura 7 ilustra um modelo de conceitual que descreve o conteúdo do framework de uma arquitetura.

Figura 7 – Framework de uma Arquitetura



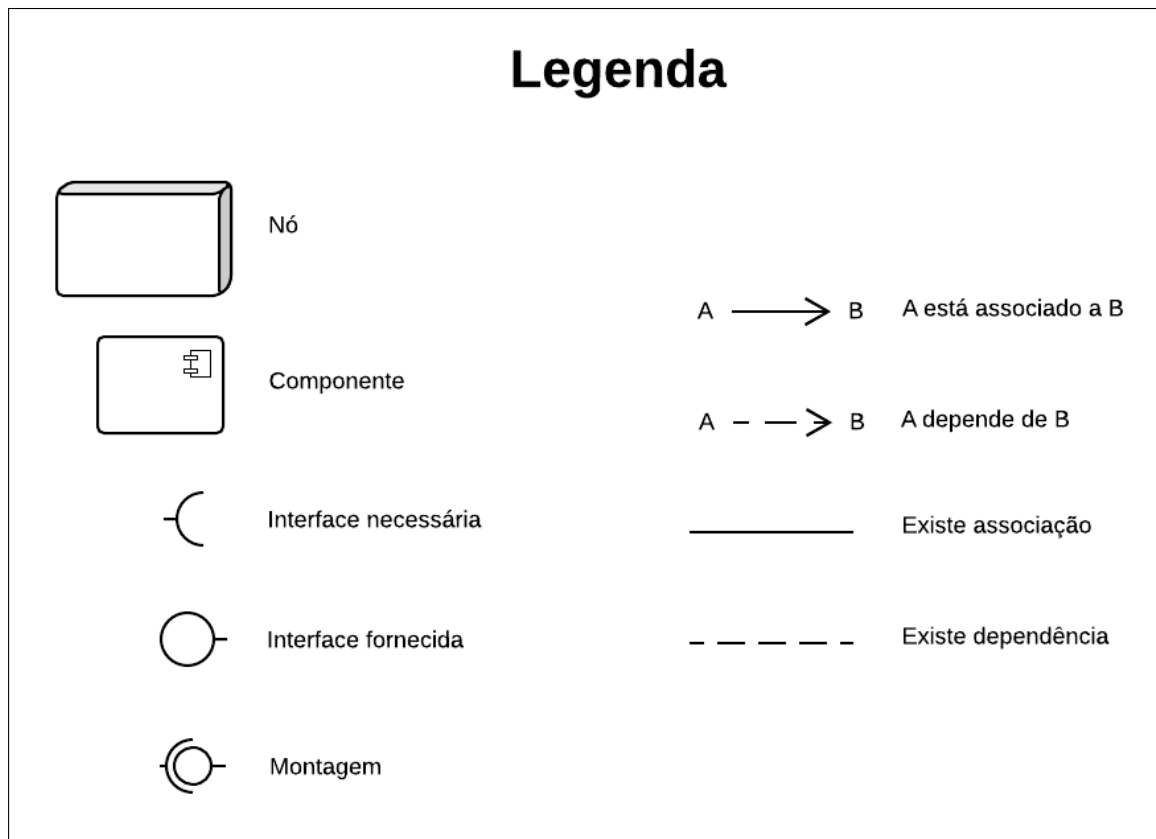
Fonte - ([International Organization Of Standardization, 2011](#))

Uma linguagem de descrição de uma arquitetura, de acordo com a norma, provê um ou mais tipos de modelos para projetar visualmente, ilustrar, interesses para o público-alvo da arquitetura. Na seção seguinte, apresentaremos um diagrama da linguagem *UML* que será utilizado no capítulo 3, com o objetivo de ilustrar o funcionamento da visão da aplicação da arquitetura.

2.6 Diagrama de Componentes

De acordo com [Guedes \(2018\)](#), o diagrama de componentes da *UML* identifica os componentes, que compõem um sistema, subsistema, ou os componentes de classes internas ou um

Figura 8 – Simbologia utilizada no diagrama de componentes



Fonte - Autores (2018)

componente individual. Neste trabalho, o diagrama de componentes foi utilizado para representar a arquitetura de dados abertos definida para a solução. Este diagrama foi escolhido, pois favorece a visualização das estruturas de software desenvolvidas, as interfaces disponibilizadas, e o relacionamento entre elas.

A Figura 8 mostra a simbologia utilizada num diagrama de componentes. A seguir, serão detalhados um pouco mais sobre cada um dos símbolos, de acordo com o manual da [Lucid Chart \(2019\)](#):

- **Nó:** O nó pode representar objetos de hardware ou de software que estão num nível acima dos componentes. Ex: Servidores, instâncias de aplicação.
- **Componente:** Uma entidade que representa um papel bem definido em um sistema. Ex: módulo de saque de um caixa eletrônico, biblioteca de conversão de moedas.
- **Interface Necessária:** Representa a necessidade de existir uma interface para estabelecer a comunicação com uma outra interface disponibilizada. Ex: Interface de autorização de saque, para uma interface de solicitação de saque, em caixas eletrônicos.
- **Interface Disponibilizada:** Representa uma interface disponibilizada para comunicação com outras interfaces. Ex: Módulo de saque para caixas eletrônicos.

- **Montagem:** Conexão entre interfaces necessária e fornecida. Ex: Autorização de saque pelo banco, e efetivação da transação.
- **Associação:** Mostra que uma parte do sistema relaciona-se com outra, mas não necessariamente estabelece relação de dependência.
- **Dependência:** Indica que uma parte do sistema além de se relacionar com outra estabelece uma relação de dependência.
- **Direção de Seta em uma Relação:** Demonstra a direção da relação estabelecida entre duas partes de um sistema.

2.7 Padrão Arquitetural REST (Representational State Transfer)

É um estilo arquitetural para a web que provê um conjunto de métodos, através do protocolo HTTP, para a comunicação entre sistemas, e os sistemas desenvolvidos com esse padrão são chamados de *RESTful*. De acordo com [Fielding \(2000\)](#), criador do padrão arquitetural em sua dissertação, *REST* é um conjunto coordenado de restrições (*constraints*) arquiteturais com o intuito de reduzir a latência e a comunicação de rede, ao mesmo tempo que maximiza a independência, e a escalabilidade da implementação dos componentes.

API's REST devem possuir algumas características, de acordo com [Fielding \(2000\)](#):

- **Arquitetura Cliente-Servidor** - Separar claramente os nós cliente dos nós servidores, em outras palavras, estabelecer uma arquitetura cliente-servidor bem definida.
- **Comunicação sem Estados** - Todas as requisições do cliente ao servidor devem conter todos os parâmetros necessários para processamento pelo servidor, e não devem se aproveitar de qualquer contexto previamente armazenado nele.
- **Cache** - Esta restrição determina que as respostas podem estar em cache no servidor, se solicitado pelo cliente, visando desempenho das próximas requisições que contenham os mesmos parâmetros.
- **Interface Uniforme** - Algumas regras de padronização devem ser estabelecidas: identificação do recurso através de *URI's*; resposta de requisições através de formatos representáveis como *JSON*, e *XML*.; mensagens auto-descritivas; e hipermídia como o motor de estados da aplicação.
- **Arquitetura em Camadas** - Definir os estágios de transição dos dados, com o objetivo de evitar que o cliente acesse diretamente camadas mais internas do servidor, sob risco de burlar regras de negócio importantes.

- **Código sob Demanda** - Esta característica é opcional e trata da execução de algumas partes do código do lado cliente, ao invés de executá-lo do lado do servidor.

Por ser implementada sob o protocolo *HTTP*, a arquitetura *REST* utiliza os seus métodos de manipulação dos recursos, conforme descrito por [Fielding \(2000\)](#). São eles:

- **POST:** Cria um novo recurso.
- **GET:** Consulta um recurso existente.
- **PUT:** Modifica um recurso existente.
- **DELETE:** Remove um recurso existente.

Arquiteturas *REST* são utilizadas para desenvolvimento de *API's* (conjunto de códigos rotinas e ferramentas para construção de software) de acesso a dados pela web. *API's RESTful* permitem, por exemplo, que cliente e servidor estejam implementados em linguagens de programação diferentes, mas ainda possam se comunicar, já que utilizam um mesmo conjunto de regras e padrões para tal.

2.8 Considerações Finais do Capítulo

Ao longo deste capítulo, foram apresentados os conceitos necessários para o entendimento do que é um portal de dados abertos e de como ele se relaciona com cidades inteligentes. Além do conceito principal, outros conceitos que permeiam o tema, como transparência governamental, tecnologias envolvidas para implementação de um portal de dados abertos também foram apresentados. Além disso, toda a teoria sobre implementação de uma arquitetura de dados abertos também foi apresentada para subsidiar o bom entendimento do tema principal, e sua relação com os produtos desenvolvidos ao longo do trabalho.

3

Arquitetura Open Data UFS

Este capítulo apresenta a arquitetura de Dados Abertos proposta, baseada na ISO/IEC/IEEE 42010:2011 ([International Organization Of Standardization, 2011](#)) da solução de dados abertos idealizada. Ela serve de base para a concepção do *PDA*, implantação do portal de dados abertos, e extração dos dados dos sistemas de informação da *UFS*, e publicação dos *datasets* no portal *Dados Abertos UFS*. Aqui descrevemos os aspectos necessários, apresentados no capítulo 2, para definir uma arquitetura. Foram utilizados diagramas de componente, apresentados também na Fundamentação Teórica, para ilustrar passo a passo o funcionamento do processo de publicação dos dados, e facilitar o seu entendimento.

3.1 Visão Geral

A arquitetura *Open Data UFS* permeou todo o processo de desenvolvimento da solução de Dados Abertos da Universidade. E, a partir dela, foi elaborado o *Plano de Dados Abertos (PDA)*, que será apresentado no capítulo 4, alinhando os objetivos gerais e de tecnologia de informação da instituição, com os interesses gerais da comunidade, como também aos aspectos legais definidos pela legislação federal. De acordo com a norma *ISO/IEC/IEEE 42010:2011* ([International Organization Of Standardization, 2011](#)), uma arquitetura de *software* deve conter as seguintes descrições: identificação de partes interessadas e interesses, visões, modelos, relações, e decisões importantes.

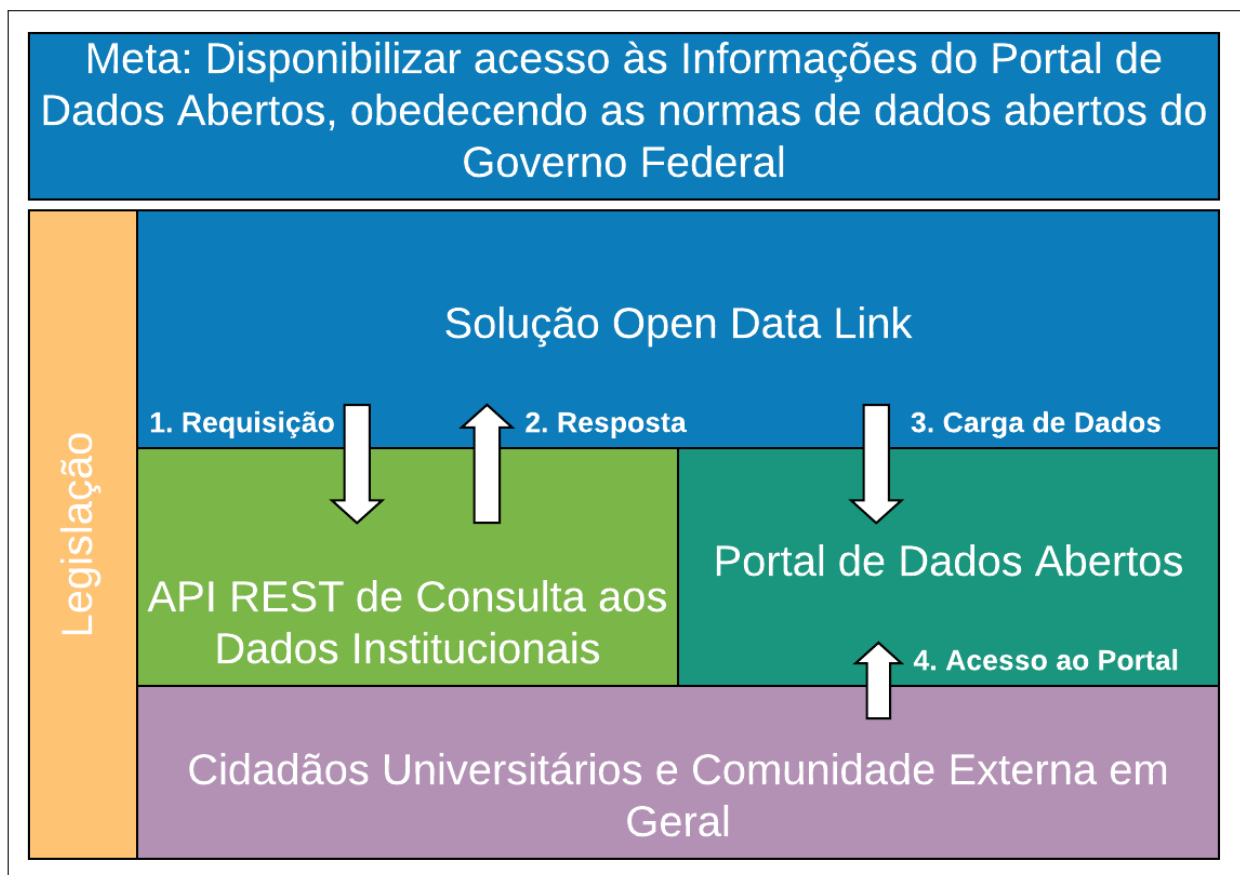
Quanto aos requisitos de *software* é importante documentá-los numa arquitetura pois podem ter impacto direto nos interesses *concerns* das partes interessadas *stakeholders*. [Sommerville \(1992\)](#) definiu os requisitos como funcionais e não-funcionais. Os requisitos funcionais são funções que o sistema ou componentes do sistema deve realizar, e os não-funcionais são descrições dos serviços que o sistema disponibiliza.

De acordo com [Ross, Weill e Robertson \(2008\)](#), uma arquitetura corporativa é a lógica

organizadora do processo de negócios e da infraestrutura de TI, que refletem os requisitos de integração e padronização do modelo operacional da empresa. Segundo os autores, este tipo de arquitetura provê uma visão duradoura dos processos, sistemas e tecnologias da companhia, de forma que seus projetos podem desenvolver capacidades, e não apenas atender necessidades momentâneas.

A norma *ISO/IEC/IEEE 42010:2011* ([International Organization Of Standardization, 2011](#)) cita uma série de frameworks corporativos como o *TOGAF* ([The Open Group, 2018](#)), e o *MODAF* ([United Kingdom Ministry of Defence, 2012](#)); frameworks de visão arquitetural, além dos dois citados anteriormente que contemplam visões arquiteturais, são citados o Modelo de Visão Arquitetural "4 + 1" ([KRUCHTEN, 1995](#)), Os 4 Modelos de visão da siemens ([HOFMEISTER; NORD; SONI, 2000](#)), dentre outros.

Figura 9 – Visão Geral da Arquitetura *Open Data UFS*



Fonte - Autores (2018)

3.2 Identificação das Partes Interessadas (*Stakeholders*)

A norma *ISO/IEC/IEEE 42010:2011* ([International Organization Of Standardization, 2011](#)) define as *partes interessadas* como indivíduo (componentes do sistema), time, organização

ou classe dos mesmos que possuem interesse no sistema. De acordo com o definido no *PDA*, as partes interessadas são:

- **Cidadãos universitários:** Este conjunto de pessoas exerce um papel importante na abertura dos dados, pois são o grupo principal a quem se destinam as informações, de forma a incentivar que os dados da universidade estejam em constante observação. Por ser bastante amplo, engloba também indivíduos de outros grupos, mas em geral são alunos, professores, funcionários, e visitantes.
- **Profissionais de Tecnologia:** Neste grupo encontram-se os desenvolvedores, analistas de tecnologia, cientistas de dados, e todos os profissionais de tecnologia interessados em utilizar os dados para desenvolver soluções tecnológicas para a população.
- **Jornalistas:** Este grupo é bastante específico pois engloba apenas os profissionais de jornalismo interessados em investigar os dados publicados, com o objetivo de prover informações subsidiadas em dados oficiais sobre a universidade à população.
- **Pesquisadores:** Grupo que pode contar com atores da própria instituição, ou pesquisadores externos, com interesse em conhecer melhor os trabalhos de pesquisa desenvolvidos na universidade.
- **Gestores da Universidade:** Grupo formado por profissionais com poder de decisão na instituição, e que necessitam de informações para exercer o seu papel. São também os interessados principais em gerir com eficiência e eficácia os recursos da instituição.
- **Governo Federal:** Importante parte interessada em promover transparência para a população em geral. Também é responsável por fiscalizar, na figura do *Controladoria Geral da União (CGU)*, se a publicação das informações está condizente com a norma legislativa.
- **Comunidade Externa:** População a nível municipal, estadual ou federal, interessada em acompanhar os dados da universidade, valendo-se da transparência das instituições.
- **AUDINT:** É o órgão técnico de Controle Interno e assessoramento ao Conselho Universitário ¹ da instituição. Este setor da *UFS* é responsável por cobrar internamente o cumprimento da legislação correlata.
- **Núcleo de Tecnologia da Informação (NTI) da UFS:** Setor de tecnologia da informação da instituição ², onde se encontram os profissionais responsáveis por manter o processo tecnológico de extração das informações, e sua devida carga no portal *Dados Abertos UFS*, assim como toda a tecnologia envolvida.

¹ AUDINT <<http://reitoria.ufs.br/pagina/6360>>

² NTI <<http://sti.ufs.br/>>

3.3 Interesses (*Concerns*)

Na norma ISO/IEC/IEEE 42010:2011 ([International Organization Of Standardization, 2011](#)), *interesses (concerns)* como os interesses relevantes para um ou mais interessados de um sistema. Os interesses influenciam em vários aspectos de como o sistema atua ao longo do seu ciclo de vida, daí vem a importância de identificá-los.

- **Transparência:** Podemos dizer que a transparência é um interesse geral no contexto do sistema, pois todos os grupos atuam de forma a buscá-la como objetivo (cidadãos universitários, comunidade externa, e governo federal, *AUDINT*, desenvolvedores do *NTI*), ou utilizam-se dela para exercer um papel na sociedade (profissionais de tecnologia, jornalistas, pesquisadores, e gestores).
- **Periodicidade das Informações:** Informações acadêmicas mudam de acordo com o período letivo; Informações financeiras podem mudar diariamente; Informações institucionais podem ser estáticas. Portanto definir uma periodicidade de atualização das informações no portal é extremamente importante.
- **Legalidade das Informações:** Além da periodicidade, o cumprimento legal no ato da publicação também é importante. Dessa forma, podemos garantir que informações sigilosas sob guarda da instituição não serão publicadas, se a instituição está respeitando a periodicidade posta no *PDA*, e se esta atendendo a demanda da comunidade.
- **Usabilidade do Portal:** Interesse dos visitantes do portal, representados pelas figuras dos *stakeholders*: cidadãos universitários, comunidade externa e pesquisadores. Este interesse é uma preocupação que o *NTI* e a *AUDINT* devem ter, para que não ocorram reduções ou impedimentos de acesso ao portal por instabilidade na infraestrutura ou baixa usabilidade da interface do portal.
- **Manutenção das Tecnologias Envolvidas:** Para os desenvolvedores de tecnologia do *NTI*, este é um interesse que tem impacto direto no fluxo de trabalho do setor. O portal *Dados Abertos UFS*, como também a solução *Open Data Link*, componentes tecnológicos imprescindíveis para o sistema, após o encerramento do projeto de implantação, entram num novo ciclo de manutenção contínua (evolutiva e eventualmente corretiva). Deste modo, os novos componentes devem integrar a agenda de manutenções do setor. Além dos novos componentes, a sua comunicação com a *API Sistemas* também deve fazer parte da lista de manutenções, com o objetivo de garantir o sucesso do fluxo de trabalho do sistema.
- **Segurança tecnológica:** Interesse relativo à segurança do parque tecnológico ao inserir o novo sistema. A equipe de desenvolvedores e analistas de segurança da informação do *NTI* são as partes diretamente interessadas, mas é um ponto que diz respeito a toda instituição, caso haja comprometimento do parque tecnológico já existente.

- **Interesses Financeiros:** O custo para a instituição de integrar o novo sistema ao seu parque tecnológico. Com relação ao aparato tecnológico, pouco ou quase nada financeiramente foi demandado, pois toda a solução foi implementada na infraestrutura que a instituição já possuía. Porém, nenhum estudo foi feito projetando para o futuro a evolução de infraestrutura necessária. Portanto esta é uma preocupação que os interessados da área de tecnologia, e gestores da instituição devem ter a partir da implantação do projeto.
- **Mobilização de Recursos Humanos:** Esta é uma preocupação que tanto os gestores, quanto o NTI e a AUDINT da instituição devem ter em mente. Pois eventuais mobilizações de recursos humanos para identificação de futuros *datasets*, como também reforço no processo de fiscalização e homologação dos dados serão necessários. Além disso, recursos humanos também podem ser necessários em etapas de todos os interesses anteriores.
- **Melhorias Sociais e Institucionais:** São interesses de eventuais melhorias sociais e institucionais caso os desenvolvedores, jornalistas, pesquisadores e gestores façam uso dos dados publicados.

3.4 Requisitos funcionais e não-funcionais

Para o portal, foi utilizado o *CKAN* como software base para personalização, pois ele atende os requisitos funcionais e não-funcionais descritos a seguir:

- Tela inicial com barra de pesquisa dos *datasets* cadastrados.
- Tela de login de usuário.
- Tela de administração do portal para funções de manutenção dos *datasets* restritas ao administrador de dados.
- Tela de listagem dos *datasets* cadastrados.
- Tela de classificação dos *datasets* cadastrados em grupos.
- Disponibilização de interface *RESTful* para carga dos *datasets*.
- Permitir personalização através de interface de arquivos de configuração.

Como o *CKAN* é um *software* desenvolvido para atender muitas necessidades, as funcionalidades que ele possuía a mais das que eram necessárias foram omitidas dos usuários visitantes, ou configuradas para visualização apenas pelo administrador do portal.

Os requisitos não funcionais do portal estão definidos a seguir:

- Ambiente com processador de no mínimo 2 núcleos.

- Ambiente com no mínimo 60Gb de espaço em disco rígido.
- Ambiente com no mínimo 4Gb de memória *RAM*.
- Tempo de resposta das requisições menor que 15s.
- Possibilidade de adição de *datasets* em grandes quantidades e automatizada.
- Permitir fácil instalação do portal com scripts de automação.

Para a solução de carga dos *datasets* no portal, o *Open Data Link*, os seguintes requisitos funcionais foram elencados:

- Componente de definição de configurações gerais como *URLs* de *download* e *upload* dos *datasets*, *tokens* de acesso a *download* pela *API Sistemas* e de *upload* no portal *Dados Abertos UFS*.
- Componente de configuração dos *datasets* como endereços *URL*, nomes, grupos, organizações, diretórios de armazenagem em *cache*.
- Função de *download* dos *datasets* através de interface *RESTful* da *API Sistemas*.
- Função de *upload* dos *datasets* através de interface *RESTful* no *CKAN*.
- Função de gravação dos *datasets* nos seguintes formatos: *JSON*, *CSV*, *XLS*, *XLSX* e *ODS*.

Quanto aos requisitos não-funcionais do *Open Data Link*, estão definidos a seguir:

- Carga de 1 arquivo de *dataset* em menos de 15s.
- Permitir reuso e extensão do código.
- Gerar os *datasets* em formatos que estejam de acordo com a classificação *5 Estrelas* (HAUSENBLAS; KIM, 2012) de abertura de dados.
- Permitir carga de novos *datasets*, a medida que forem identificados.

3.5 Visões e Pontos de Vista (*Views e Viewpoints*)

De acordo com definição da norma *ISO/IEC 10746-1:1998* (International Organization Of Standardization, 1998), um ponto de vista é uma abstração que reproduz uma especificação de todo o sistema, relacionada a um conjunto de interesses (*concerns*). Dessa forma, a partir da identificação dos interessados, e seus respectivos interesses, foi possível classificá-los em conjuntos que definirão as visões, e consequentemente os pontos de vista da arquitetura *Open Data UFS*.

Segundo a *ISO/IEC/IEEE 42010:2011* ([International Organization Of Standardization, 2011](#)), o uso de *frameworks* de arquiteturas incluem, mas não estão limitados a, criar descrições arquiteturais, desenvolver ferramentas de modelagem, e métodos arquiteturais. Dessa forma, a visão de arquitetura corporativa e o conhecimento dos frameworks citados pela norma foram importantes para a descrição das visões e pontos de vista da arquitetura, baseados no *TOGAF* ([The Open Group, 2018](#)). O anexo B da norma *ISO/IEC/IEEE 42010:2011* ([International Organization Of Standardization, 2011](#)) traz uma ficha exemplo para descrição das visões, cujos atributos foram utilizados nas subseções a seguir.

É importante observar que o *TOGAF* ([The Open Group, 2018](#)) nomeia suas visões como arquiteturas, muito embora a norma *ISO/IEC/IEEE 42010:2011* ([International Organization Of Standardization, 2011](#)) afirme que não se deve utilizar termos como "arquitetura de negócio", "arquitetura de dados", pois entendem que a arquitetura é uma concepção mais ampla das propriedades de um sistema, e que é melhor entendida como as visões sob a ótica dos interessados na arquitetura. Dito isto, aqui determinamos que as arquiteturas extraídas do *TOGAF* ([The Open Group, 2018](#)) serão chamadas de visões.

3.5.1 Ponto de Vista do Negócio

Este ponto de vista abrange o negócio principal referente ao sistema, que neste contexto é a transparência. Como afirmamos anteriormente no item de transparência, na seção 3.3, todos os interessados têm como meta promover transparência, ou se utilizam da dela para exercer suas atividades.

- **Interesses:** Transparência
- **Partes Interessadas:** Todos os *stakeholders* são interessados neste ponto de vista, pois é o negócio principal do sistema.
- **Modelos Sugeridos:** Diagrama *BPMN*.
- **Regras de Correspondência:** A visão do negócio é o ponto de partida para todas as outras visões da arquitetura.
- **Operações:** Métodos de criação (meios pelos quais as visões são preparadas), e métodos de análise (utilizados para checar, justificar, transformar, prever, aplicar e avaliar resultados desta visão).
- **Fontes:** 1. Adaptações do *TOGAF* ([The Open Group, 2018](#)); 2. [International Organization Of Standardization \(2011\)](#).

Sob o ponto de vista do negócio, para alcançar a transparência via plano de dados abertos, foi necessário aliar os interesses gerais com os interesses tecnológicos da instituição relacionando

o *Plano de Desenvolvimento Institucional (PDI)* ao *Planejamento Estratégico de tecnologia de Informação e Comunicação* da UFS, garantindo, dessa maneira, que há interesses em comum nestes dois planejamentos que justificam a política de transparência e dados abertos. Não foram implementados diagramas nesta etapa, pois o trabalho foi apenas de análise de documentos.

3.5.2 Ponto de Vista dos Dados

Ponto de vista que define a seleção e fluxo dos dados até a sua persistência no portal.

- **Interesses:** Periodicidade e legalidade das informações, melhorias sociais e institucionais, mobilização de recursos humanos.
- **Partes Interessadas:** Todos os *stakeholders* do projeto.
- **Modelos Sugeridos:** Diagrama lógico, dicionário de dados, modelo entidade-relacionamento dos datasets.
- **Regras de Correspondência:** A visão dos dados são publicados no portal no formato de *datasets*, portanto impulsionam a visão da aplicação.
- **Operações:** Métodos interpretativos (meios pelos quais as visões devem ser entendidas pelo leitor ou pelas partes interessadas do projeto).
- **Fontes:** 1. Adaptações do *TOGAF* ([The Open Group, 2018](#)); 2. [International Organization Of Standardization \(2011\)](#).

Nesta etapa, a seleção dos dados ocorreu de acordo com os seguintes critérios, ordenados por seus respectivos pesos:

- **Necessidade de cumprimento da legislação (peso 4):** Com relação a este item, nenhum dispositivo legal impõe qual conjunto de dados deve ser publicado.
- **Consistência de informação (peso 3):** Só devem ser publicados os *datasets* com consistência de informação, com o objetivo de não motivar qualquer questionamento sobre a idoneidade e a lisura do processo de abertura dos dados.
- **Relevância para a comunidade (peso 2):** Consoante às demandas sociais elencadas na Seção 4.2, devem ser publicados os *datasets* com significativa relevância e interesse à comunidade. Como exemplo, pode-se elencar os *datasets* do grupo financeiro como empenhos emitidos pela instituição, diárias e passagens, contratos e convênios, atividades de extensão e projetos de pesquisa.

- **Facilidade de acesso à informação para consulta (peso 1):** Dentre os conjuntos de dados disponíveis na instituição serão extraídos na primeira iteração aqueles que forem de fácil acesso para extração, tiverem sua regra de consulta bem definida, e puderem ser definidos para consulta na *API Sistemas* do *NTI*. Este item invalida completamente a prioridade dos itens anteriores se a sua viabilidade de acesso for nula.

Quando dois critérios geram conflitos de prioridade (fácil acesso porém informação inconsistente, ou difícil acesso mas informação consistente), prevalece o de maior peso. É importante estabelecer que quando a viabilidade de obter o *dataset* for nula, o caso deve ser discutido entre as partes interessadas.

Também é importante definir filtros de impedimento da publicação das informações como:

- Não divulgar dados pessoais dos indivíduos cadastrados nas bases da instituição.
- Não divulgar dados sigilosos da instituição.

3.5.3 Ponto de Vista da Aplicação

Define as aplicações, e suas relações com o negócio do sistema, além de compilar, organizar e utilizar as tecnologias definidas na arquitetura.

- **Interesses:** Usabilidade do portal, periodicidade das informações, manutenção das tecnologias envolvidas, segurança tecnológica.
- **Partes Interessadas:** O público visitante do portal (cidadãos universitários, profissionais de tecnologia, jornalistas, pesquisadores), e o *NTI*.
- **Modelos Sugeridos:** Dicionário de dados para o público visitante, diagrama de componentes para o *NTI*.
- **Regras de Correspondência:** A visão da aplicação determina as tecnologias utilizadas na visão de tecnologias.
- **Operações:** Métodos interpretativos (São os meios pelos quais as visões devem ser entendidas pelo leitor ou pelas partes interessadas do projeto).
- **Fontes:** 1. Adaptações do *TOGAF* ([The Open Group, 2018](#)); 2. [International Organization Of Standardization \(2011\)](#).

Para esclarecer este ponto de vista, a seção 3.9 apresenta uma série de diagramas de componentes com a representação em detalhes dos componentes da solução de extração e publicação dos dados, o *Open Data Link*, e seu funcionamento.

3.5.4 Ponto de Vista da Tecnologia

Determina as tecnologias envolvidas no ciclo de vida do sistema.

- **Interesses:** Manutenção das tecnologias envolvidas, segurança tecnológica, financeiros.
- **Partes Interessadas:** Gestores da universidade, e o *NTI*.
- **Modelos Sugeridos:** Diagrama de componentes para o *NTI*, e implantação para os gestores e o *NTI*.
- **Regras de Correspondência:** As visões anteriores influenciam diretamente nas tecnologias envolvidas no ciclo de vida do sistema.
- **Operações:** Métodos interpretativos (São os meios pelos quais as visões devem ser entendidas pelo leitor ou pelas partes interessadas do projeto).
- **Fontes:** 1. Adaptações do *TOGAF* ([The Open Group, 2018](#)); 2. [International Organization Of Standardization \(2011\)](#).

3.6 Tecnologias Envolvidas na Implantação do Portal Dados Abertos UFS

O *CKAN* conta com tecnologias auxiliares para o seu funcionamento adequado, que estão descritas a seguir:

- **jetty³:** Servidor *HTTP Java* e *container servlet Java*.
- **solr⁴:** Plataforma de busca textual desenvolvida em *Java*.
- **nginx⁵:** Servidor *web*, que funciona como servidor de cache, balanceador de carga, e proxy reverso. O seu uso é opcional e não foi utilizado no projeto.
- **PostgreSQL⁶:** Banco de dados relacional de código aberto. Utilizado para persistir as informações do *CKAN*.
- **Jinja⁷:** Linguagem de templates para *Python*. Utilizada para escrever o código do *front-end* do *CKAN*.

³ *jetty* - <<https://www.eclipse.org/jetty>>

⁴ *solr* - <<https://lucene.apache.org/solr>>

⁵ *nginx* - <<https://www.nginx.com>>

⁶ *PostgreSQL* - <<https://www.postgresql.org>>

⁷ *Jinja* - <<http://jinja.pocoo.org/>>

- **Pylons**⁸: *Framework web* para *Python*. Auxilia no desenvolvimento do arcabouço da aplicação.
- **SQLAlchemy**⁹: Biblioteca de mapeamento objeto-relacional para *Python*. Utilizada para realizar mapeamento objeto-relacional e facilitar o acesso aos dados das entidades, descritas no código fonte, no banco de dados.
- **Python**¹⁰: Linguagem de programação de alto-nível, interpretada, multiparadigma e de propósito geral. utilizada para escrever o código do *back-end* do *CKAN*.
- **Javascript**¹¹: Linguagem de programação de alto-nível, interpretada, multiparadigma, e uma das três tecnologias mais utilizadas na *web*. Utilizada para escrever o código do *front-end* do *CKAN*.
- **HTML**¹²: Linguagem de marcação de texto para a *web*. Assim como o *Javascript* é uma das três tecnologias mais utilizadas na *web*. Utilizada para escrever o código do *front-end* do *CKAN*.
- **msgfmt** - Aplicativo de compilação de um catálogo de mensagens para um formato binário aceito pelo *CKAN*.

3.7 Personalização do CKAN para o Portal Dados Abertos UFS

O Portal *Dados Abertos UFS* foi desenvolvido a partir da ferramenta *CKAN*, que é um software aberto para publicação de *datasets*. O *CKAN* é uma ferramenta extensa, e tem várias funcionalidades. Portanto, após a instalação e configuração, no primeiro acesso, o administrador poderá perceber que ele vem com todas as funcionalidades completamente liberadas. A exemplo disto, podemos citar o cadastro de usuários, que deveria ser restrito a poucos, vem completamente aberto possibilitando a qualquer usuário visitante se cadastrar no portal e receber um *token* de acesso para publicação de dados.

O usuário administrador pode acessar todas as funcionalidades, e apenas o usuário visitante deve ter seu acesso restrito. Para tal, foram inseridas no código do *CKAN* marcações (*tags*) do *jinja*, no código *HTML*, com testes lógicos do tipo do usuário, que liberava a função caso o usuário fosse administrador. Dessa forma, o primeiro desafio na personalização do *CKAN* foi justamente limitar as suas funcionalidades para restringir o acesso dos usuários visitantes às seguintes funcionalidades:

⁸ *Pylons* - <<https://pylonsproject.org/>>

⁹ *SQLAlchemy* - <<https://www.sqlalchemy.org>>

¹⁰ *Python* - <<https://www.python.org>>

¹¹ *Javascript* - <<https://www.javascript.com>>

¹² *HTML* - <<https://www.w3.org/TR/html52>>

- **Pesquisar datasets**
- **Listar e realizar download de datasets**
- **Listar grupos e datasets de grupos**
- **Acessar página explicativa sobre o portal**

Todas as outras funcionalidades estão liberadas para o administrador. Em seguida, foi necessário corrigir o texto da página, pois a tradução sugerida para português do Brasil do *CKAN* estava bastante literal, e alguns termos da *web* são mais conhecidos no inglês, como *Download*. O arquivo *ckan.po* no diretório *ckan/ckan/i18n/pt_BR/LC_MESSAGES/* do projeto do *CKAN* contém todas as mensagens originais do sistema com a sua respectiva tradução para o português do Brasil. É importante notar que o diretório *ckan/ckan/i18n/*, do projeto do *CKAN*, contém os subdiretórios correspondentes às linguagens também de outros países. Após a edição, o arquivo deve ser compilado utilizando o aplicativo *msgfmt* executado a partir do terminal, da seguinte forma:

```
$ msgfmt ckan.po -o ckan.mo
```

Com o objetivo de valorizar e refletir as cores da instituição no portal, foi feito um trabalho de modificação das suas cores no arquivo *CSS* do projeto do *CKAN*. Para tal, foi executado o seguinte procedimento:

1. Cópia do arquivo *main.css* disponível no diretório. *ckan/ckan/public/base/css/* do projeto do *CKAN*, para o arquivo *ufs.css*.
2. O arquivo *ufs.css* foi editado de acordo com decisão em conjunto com a equipe de superintendentes do *NTI*, com o objetivo de contemplar as cores oficiais da instituição.
3. Adição de referência ao novo arquivo *CSS* com o estilo personalizado da *UFS* ao código fonte do *CKAN*, no módulo *ckan/ckan/controller/admin.py*, conforme o código abaixo:

```
styles = [{ 'text': 'Default', 'value': '/base/css/main.css'},  
          { 'text': 'Red', 'value': '/base/css/red.css'},  
          { 'text': 'Green', 'value': '/base/css/green.css'},  
          { 'text': 'Maroon', 'value': '/base/css/maroon.css'},  
          { 'text': 'Fuchsia', 'value': '/base/css/fuchsia.css'},  
          { 'text': 'UFS', 'value': '/base/css/ufs.css'}]
```

4. A variável *ckan.main_css* presente no arquivo *ckan/ckan/config/deployment.ini_tmpl*, que contém a configuração principal do *CKAN*, foi editada para receber o valor referente ao novo estilo personalizado contendo as cores da instituição, conforme o código abaixo:

```
ckan.main_css = /base/css/ufs.css
```

Por fim, era necessário que ao contruir o portal algumas configurações já estivessem pré-definidas, dessa forma, as imagens do logotipo e do ícone foram devidamente adicionadas ao diretório *ckan/ckan/public/base/images*, que contém as imagens do *CKAN*, e em seguida foram pré-configuradas as variáveis a seguir, no arquivo *ckan/ckan/config/deployment.ini_tmpl*: *ckan.main_css*, *ckan.site_title*, *ckan.site_logo*, *ckan.site_description*, *ckan.site_about*, *ckan.site_intro_text*, *ckan.favicon*, *ckan.homepage_style*, *ckan.locale_order*, e *ckan.datapusher.formats*.

3.8 Instalação do Portal Dados Abertos UFS no Servidor

Existem três formas de instalar o *CKAN* no servidor: compilando o seu código fonte; instalando os pacotes pré-compilados para o sistema operacional *Ubuntu*¹³; ou instalando a partir do *Docker*¹⁴.

A instalação escolhida para o portal foi utilizando o *Docker*, pois a imagem da instalação já vem pronta com todas as suas dependências satisfeitas e pré-configuradas. A instalação através do pacote para *Ubuntu* foi realizada em momento anterior, porém o processo foi mais complexo, e foi necessário instalar e configurar as tecnologias listadas na seção 3.6, para o correto funcionamento do *CKAN*. A instalação via *Docker*, por outro lado, foi mais simples e ocorreu sem empecilhos, bastando segui-la criteriosamente. Mas, posteriormente, foi desenvolvido, neste trabalho, um script de instalação do *doker* com o *CKAN* já configurado no servidor, facilitando bastante o processo de instalação e configuração do portal.

O manual utilizado para instalação via *Docker* encontra-se disponível na [documentação oficial do CKAN](#)¹⁵.

3.9 Representação em Componentes do Procedimento de Publicação das Informações no Portal

Nesta Seção, serão elucidados mais detalhes sobre a Arquitetura *Open Data UFS*, apresentada na Seção anterior, e seu funcionamento, explicando como a estrutura da solução foi desenvolvida baseada em componentes substituíveis. Esta é a visão da aplicação representada passo a passo nas subseções seguintes, num diagrama de componentes.

Para coordenar o processo de extração e publicação dos dados, foi desenvolvida uma solução batizada de Solução *Open Data Link*, escrita na linguagem de programação *Python*¹⁶, pois todas as bibliotecas de manipulação de datasets nos formatos *JSON*, *CSV*, *XLS*, *XLSX*, e

¹³ *Ubuntu* - <<https://www.ubuntu.com>>

¹⁴ *Docker* - <<https://www.docker.com>>

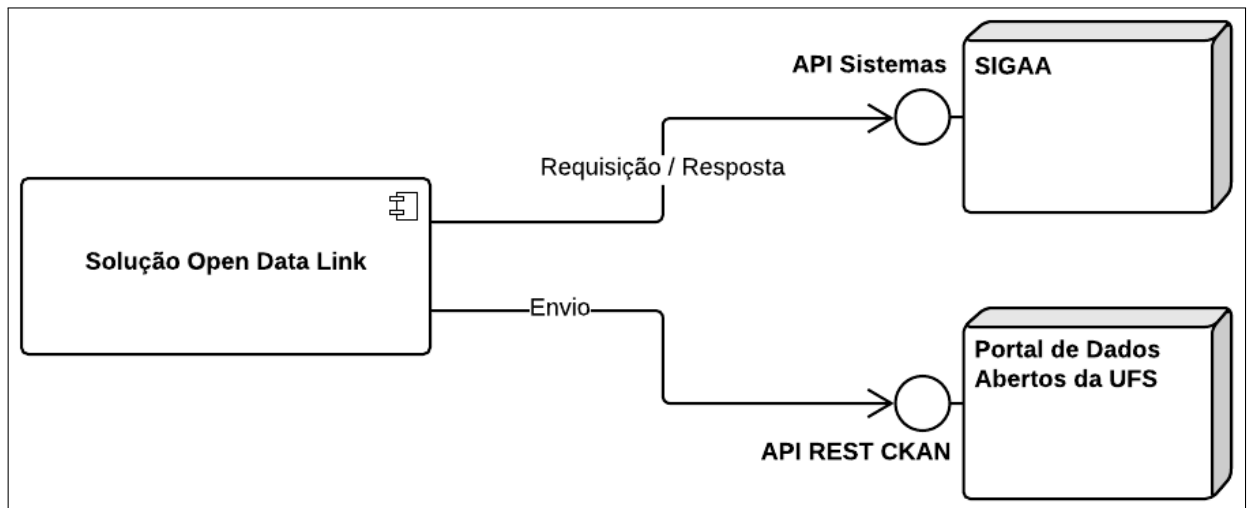
¹⁵ Documentação Oficial do *CKAN* - <<https://docs.ckan.org/en/latest/maintaining/installing/install-from-docker-compose.html>>

¹⁶ *Python* - <<https://www.python.org>>

ODS, necessárias ao processo, puderam ser facilmente encontradas. Além disso, a equipe já possuía experiência e domínio da linguagem, o que acabou forçando a escolha inicial.

A Figura 10 mostra a visão geral do procedimento de carga dos dados no portal. Nela podemos perceber que o componente *Solução Open Data Link* acessa as interfaces *API Sistemas*, para solicitar os dados, e a *API REST CKAN*, para publicação no portal *Dados Abertos UFS*

Figura 10 – Visão Geral do Procedimento de Carga dos Dados no Portal

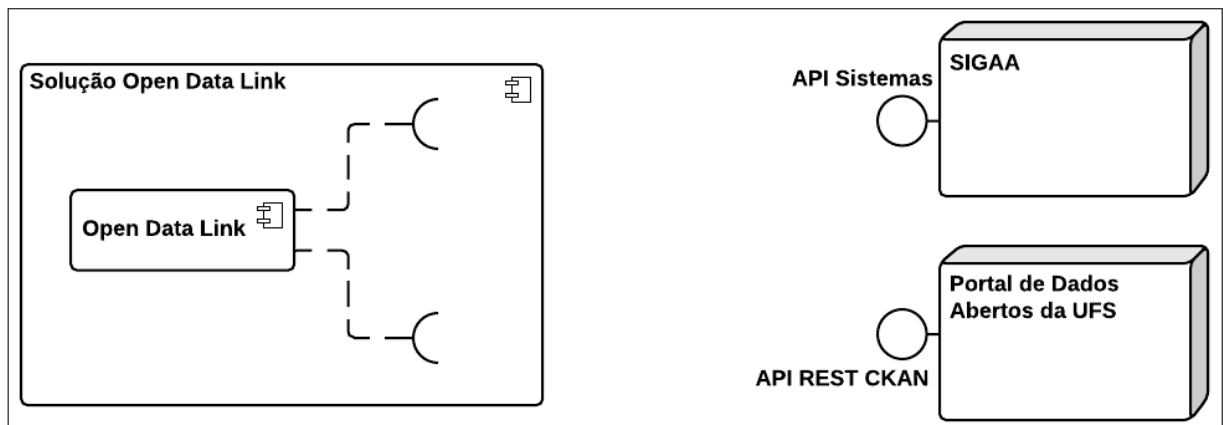


Fonte - Autores (2018)

3.9.1 As Duas Interfaces Necessárias

Na Figura 11, podemos observar os componentes da Arquitetura *Open Data UFS* sem as suas conexões. Seria necessário desenvolver as duas interfaces de acesso às interfaces disponibilizadas, *API Sistemas* e *API REST CKAN*.

Figura 11 – As Duas Interfaces Necessárias



Fonte - Autores (2018)

A interface disponibilizada para *download* das informações da instituição, disponibilizada pelo *NTI* (Núcleo de Tecnologia da Informação), é chamada de *API Sistemas*. Toda e qualquer extração de dados deve ser feita através dela pois, dessa forma, o *NTI* pode garantir que as regras intrínsecas ao negócio estão sendo aplicadas, além de evitar duplicação de processos e códigos que realizam as mesmas tarefas. A *API Sistemas* foi desenvolvida na linguagem de programação *Java*¹⁷, sob o padrão arquitetural *RESTful*. Sendo a *API Sistemas* uma interface desenvolvida pela própria instituição, a probabilidade de encontrar uma biblioteca de acesso a ela (ou que pelo menos que fosse reutilizável e reduzisse o trabalho de implementação), na *web* seria mínima portanto, o seu desenvolvimento durante o trabalho foi necessário.

A biblioteca desenvolvida para acesso à interface *API Sistemas* foi batizada com o nome de *lib_apisistemas*. O desenvolvimento dessa biblioteca contribui com outros desenvolvedores que desejem acesso aos dados da instituição, e que, a partir do seu desenvolvimento, poderão reutilizá-la nas suas aplicações.

A interface disponibilizada para *upload* de informações pelo *CKAN* também segue o mesmo padrão arquitetural *RESTful* da *API Sistemas*. A implementação dessa biblioteca de acesso à *API REST* do *CKAN* foi batizada de *lib_ckanUFS*. Ela implementa as funções de *upload* de *datasets*, e consulta destes para evitar reposição de informações. A *lib_ckanUFS* foi disponibilizada em um repositório no *GitHub*¹⁸, no endereço <https://github.com/vatbarreto/opendata_link>, assim como o projeto de personalização do *CKAN*, e se algum desenvolvedor desejar, pode bifurcar o projeto e continuar o seu desenvolvimento.

3.9.2 A Montagem dos Componentes

A Figura 12 mostra a montagem dos componentes desenvolvidos necessários ao acesso às *API's REST* disponibilizadas.

A montagem da interface necessária *lib_apisistemas* com a interface disponibilizada *API Sistemas* permite que o componente *Open Data Link* requisiute informações, que por sua vez serão respondidas com um conjunto de dados no format *JSON*. De maneira análoga, a montagem da interface necessária *lib_ckanUFS* com a interface disponível *API REST CKAN* permite que o componente *Open Data Link* envie os *datasets* para publicação no Portal *Dados Abertos UFS*.

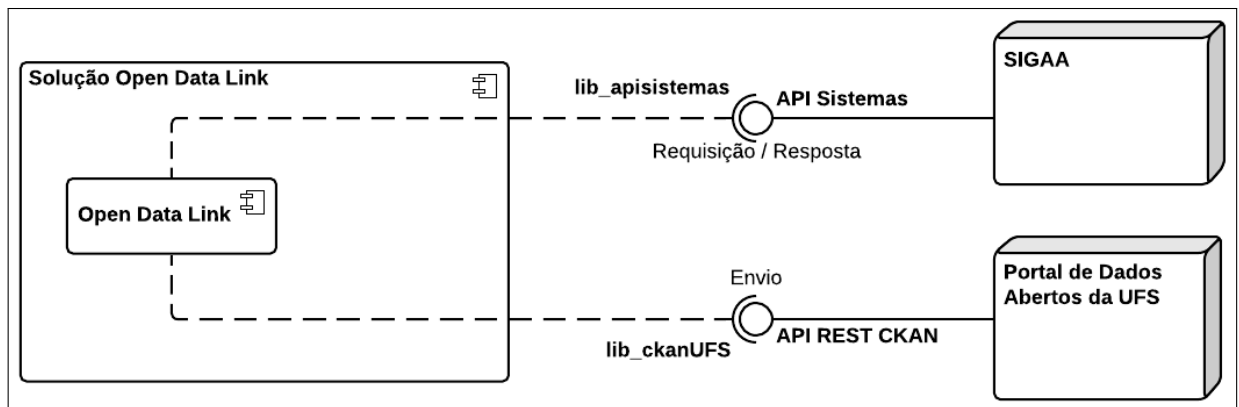
3.9.3 Representação das Interfaces Necessárias como Componentes

A Figura 13 mostra o mesmo esquema da Figura 12, porém, representando as interfaces necessárias agora como componentes da *Solução Open Data Link*. Essa Figura é importante para dar ênfase à questão do reuso, ou substituição, dessas duas interfaces.

¹⁷ *Java* - <<https://www.oracle.com/technetwork/java/index.html>>

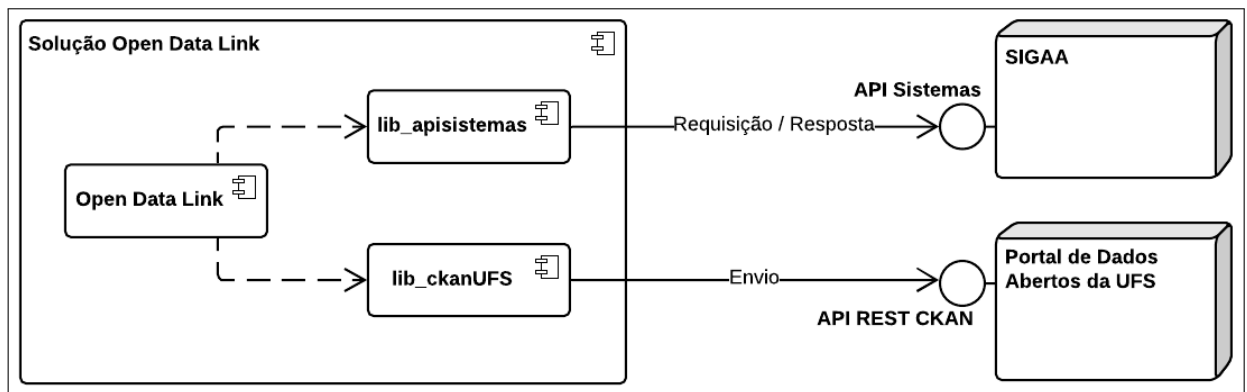
¹⁸ *GitHub* - <<https://github.com>>

Figura 12 – A Montagem dos Componentes



Fonte - Autores (2018)

Figura 13 – Representação das Interfaces Necessárias como Componentes



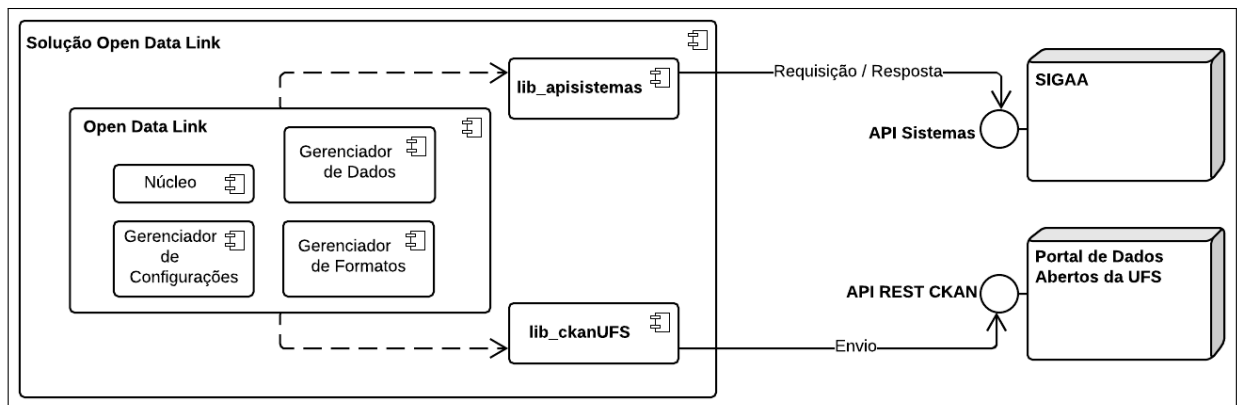
Fonte - Autores (2018)

3.9.4 Detalhes do Componente Open Data Link

Por fim, a Figura 14 tece mais detalhes sobre os componentes internos do componente principal do processo de extração e publicação dos dados, o *Open Data Link*.

O componente *Open Data Link* possui um núcleo de execução principal, e três gerenciadores: de configurações, de dados, e formatos. A seguir, há uma breve explicação sobre cada um dos componentes:

- **Núcleo:** É responsável pela execução principal do componente *Open Data Link*. É ele que instancia os gerenciadores quando necessário, solicita o comando de *download* e *upload* dos dados ao *Gerenciador de Dados*, solicita o comando de conversão dos dados, ao *Gerenciador de Formatos*, nos formatos de *datasets* necessários.
- **Gerenciador de Configurações:** Este gerenciador é responsável por armazenar e gravar, em arquivo, as configurações necessárias para o funcionamento da solução. Dentre as configurações gerenciadas por este componente estão: validade do *token* de acesso à *API*

Figura 14 – Detalhes do Componente *Open Data Link*

Fonte - Autores (2018)

Sistemas; os dois *tokens* de acesso às *API's* de *download* e *upload* das informações, os endereços de *download* e *upload* dos datasets.

- **Gerenciador de Dados:** Este gerenciador possui três tarefas: demandar o *download* das informações e armazená-las em um formato *JSON* temporário, eliminar as informações desnecessárias dos datasets, e realizar o *upload* dos *datasets* no Portal *Dados Abertos UFS*.
- **Gerenciador de Formatos:** Este gerenciador é responsável por converter os dados dos *datasets* nos formatos *JSON*, *CSV*, *ODS*, *XLS* e *XLSX*.

3.10 Implementação da Solução Open Data Link

Em princípio, uma ferramenta ETL seria utilizada para extrair os dados da base de dados do SIGAA e dos outros sistemas disponíveis (se houvesse). De acordo com Vassiliadis (2009), A sigla *ETL* significa *Extract, Transform, Load* (Extração, Transformação, Carregamento) e é o *software* que extrai, trata o resultado da extração, e em seguida carrega os dados num repositório.

Porém, o *NTI* da UFS determinou que qualquer acesso aos dados de sistemas de informação e comunicação da universidade deveria ser feito através de uma *API REST*, batizada de *API Sistemas*, que estava sendo desenvolvida pelo setor, e qualquer acesso ou instalação de aplicativos para extração de dados estava vetada. De fato, com o objetivo de resguardar a segurança das informações e da estrutura tecnológica da UFS, o acesso direto aos dados pode gerar brechas de segurança.

Dessa forma, o único acesso aos dados para extração deveria ser feito através da *API Sistemas* e uma ferramenta deveria ser desenvolvida para extração dos dados do SIGAA e carregamento no Portal *Dados Abertos UFS*.

A ferramenta desenvolvida para intermediar a relação entre a interface fornecida pelo NTI e o Portal *Dados Abertos UFS* foi batizada de *Open Data Link*. Como demonstrado no capítulo 3, na arquitetura de dados abertos da solução, a ferramenta é composta de 3 partes essenciais:

- **Open Data Link:** Parte principal da solução, que coordena todas as outras ações da ferramenta, como por exemplo, instanciar as outras bibliotecas, demandar o download, gravação, e upload dos datasets nos formatos necessários.
- **lib_apisistemas:** Biblioteca que serve de interface de acesso à API Sistemas. Esta biblioteca obrigatoriamente seria desenvolvida, já que não seria possível o uso de uma ferramenta ETL para extração e tratamento dos dados selecionados para publicação no Portal. Esta biblioteca, pode ser reutilizada por qualquer desenvolvedor que deseje acessar a *API Sistemas*, e tenha algum domínio da linguagem *Python*, e ela já possui algumas entidades implementadas (utilizadas no processo de publicação de dados no Portal *Dados Abertos UFS*).
- **lib_ckanUFS:** Biblioteca que serve de interface de carregamento dos *datasets* no Portal *Dados Abertos UFS*, através da *API REST* do *CKAN*.

Buscando agilidade no desenvolvimento, e o maior reuso de código possível, a linguagem escolhida para desenvolvimento da solução foi o *Python*¹⁹. A linguagem *Python* já possuía as bibliotecas de manipulação de *datasets* nos formatos *JSON*, *CSV*, *XLS*, *XLSX* e *ODS* necessárias ao projeto. Além disso, a equipe de desenvolvimento tinha domínio da linguagem, portanto ela foi a escolha mais natural.

As bibliotecas utilizadas no desenvolvimento da solução foram:

- **json:** A biblioteca lida com a conversão de *datasets* para o formato *JSON*, e deste formato para outros, como um dicionário de dados na linguagem *python*.
- **pandas:** A biblioteca *Pandas*²⁰ foi desenvolvida para trabalhos com análise de dados utilizando planilhas de texto, permite manipular datasets nos formatos *CSV* (padrão aberto não-proprietário), e *XLS* e *XLSX* da *Microsoft*²¹.
- **pyexcel:** A biblioteca lida também com datasets nos formatos *CSV* (padrão aberto não-proprietário), e *XLS* e *XLSX* da *Microsoft*²¹, como também o formato *ODS* do *Open Office*²².

¹⁹ Python: <<https://www.python.org>>

²⁰ pandas: <<https://pandas.pydata.org/>>

²¹ Microsoft: <<https://www.microsoft.com>>

²² Open Office: <<https://www.openoffice.org>>

- **pyexcel-ods3:** Desenvolvida na versão 3.0 do *python*, a biblioteca lida com datasets no formato *ODS* do *Open Office*²².
- **xlwt:** A biblioteca satisfaz as dependências da *pyexcel-ods3*, e tem funções para gerar arquivos no antigo formato *XLS* da *Microsoft*²¹.
- **xlswriter:** A biblioteca satisfaz as dependências da *pyexcel-ods3*, e implementa as funções de escrita no formato *XLSX* da *Microsoft*²¹.

3.11 Recursos da API Sistemas

A *API Sistemas* é a interface de consulta de dados dos sistemas da UFS, como o *SIGAA* (Sistema Integrado de Gestão de Atividades Acadêmicas) e o *SIPAC* (Sistema Integrado de Patrimônio Administração e Contratos). Foi desenvolvida pelo *NTI* (Núcleo de Tecnologia da Informação) da UFS, com o propósito de disponibilizar uma interface padronizada de consulta das informações de seus sistemas e, dessa forma, incentivando o desenvolvimento de aplicativos para a universidade, por desenvolvedores externos ao setor.

A *API Sistemas* foi desenvolvida para ser uma interface *RESTful* de acesso aos dados dos sistemas *SIG* (como são "apelidados" o conjunto destes sistemas de gestão na *UFS*), e qualquer acesso aos dados da instituição deve ser feito através dela.

Inicialmente, a *API Sistemas* tinha poucos recursos de consulta implementados, que servissem para o propósito do Portal *Dados Abertos UFS*. Para agravar a situação, a equipe de desenvolvimento do *NTI* é pequena para atender toda a demanda de desenvolvimento e manutenção de *software* da universidade, e não poderia desenvolver, em tempo hábil necessário à conclusão deste trabalho, todos os *datasets* necessários. Diante do cenário formado, tornou-se necessário implementar os recursos a seguir para extração de dados pro portal *Dados Abertos UFS*:

- **Estagiários:** Recurso de consulta aos dados de estágio cadastrados e supervisionados pela instituição. Não foi incluído na versão final de homologação, por falta de consistência nos relacionamentos entre as entidades. Atualmente está marcado para ser analisado e republicado.
- **Discentes:** Recurso de consulta dos discentes por ano de ingresso da instituição. Já existiam consultas de discentes, porém internas para o funcionamento das aplicações *SIG*. O trabalho atuou em definir uma estrutura que pudesse ser retornada através da *API*.
- **Empenho:** Recurso de consulta aos dados financeiros de empenhos da instituição.
- **Liquidação:** Recurso de consulta aos dados financeiros de liquidações de empenhos da instituição.

- **Ordens Bancárias:** Recurso de consulta aos dados financeiros de ordens bancárias emitidas pela instituição.
- **Projetos de Pesquisa:** Recurso de consulta aos projetos de pesquisa da instituição. Atualmente está marcado para ser homologado e republicado.
- **Atividades de Extensão:** Recurso de consulta às atividades de extensão da instituição. Atualmente está marcado para ser homologado e republicado.

3.12 Desafios Superados e Sugestões

Alguns recursos apesar de terem sido implementados não foram utilizados para alimentar o portal, pois ainda não estavam consistentes o suficiente. Os motivos foram: funcionalidades que não estavam sendo alimentadas pelos usuários da instituição; baixa disponibilidade dos analistas capacitados para orientar o entendimento de regras de negócio necessárias, consequência da alta demanda à em relação ao quantitativo da equipe. Por este motivo, foram marcados para revisão e futura implantação os *datasets*: Liquidação de Empenho; Ordens Bancárias; Contratos; Convênios; Diárias; Passagens; Atividades de Extensão; Projetos de Pesquisa.

A impossibilidade de acesso à base de dados dos sistemas da *UFS*, fora do ambiente do *NTI*, limitou o ritmo de trabalho em relação ao necessário, em determinados momentos. Para suprir esta necessidade, ficou registrada pela equipe de desenvolvimento a solicitação para ser disponibilizado um ambiente de testes público da base de dados, que possa ser acessado externamente, com dados embaralhados ou fictícios. Outra alternativa seria disponibilizar uma cópia da base com as mesmas características (dados embaralhados ou fictícios) para download.

Disponibilização do código da *API Sistemas* para download, pois a *UFS* não é a única instituição a utilizar os sistemas *SIGs* desenvolvidos pela *UFRN*. Portanto, a liberação do código, e a disponibilidade de uma base de testes para download, potencializa o número de desenvolvedores da própria instituição, fora do *NTI*, como também externos à *UFS*, habilitados a trabalhar neste projeto.

3.13 Considerações Finais do Capítulo

Este capítulo propôs a arquitetura *Open Data UFS* identificando as partes interessadas e seus respectivos interesses, requisitos funcionais e não-funcionais, visões, modelos, regras de correspondência e operações da arquitetura. Foi apresentado também o funcionamento da visão da aplicação utilizando diagramas de componentes da *UML*. O desenvolvimento do portal *Dados Abertos UFS*, o desenvolvimento do componente *Open Data Link*, os recursos desenvolvidos para a *API Sistemas*, os desafios encontrados durante o projeto e por fim foram elencadas algumas sugestões.

4

Plano de Dados Abertos (PDA) e Portal Dados Abertos UFS

Este capítulo apresenta o planejamento para abertura dos dados da instituição, denominado de Plano de Dados Abertos e, além dele, também será apresentado o portal *Dados Abertos UFS*.

Para lograr êxito ao elaborar um Plano de Dados Abertos (*PDA*), era necessário ampará-lo em quatro pilares:

- **Embasamento Legal:** os dispositivos legais que amparam o processo de abertura de dados das instituições públicas.
- **Motivação Social:** Motivos sociais que levam uma instituição a abrir os seus dados à comunidade.
- **Alinhamento ao Planejamento Estratégico:** Conciliar o projeto de abertura dos dados ao planejamento traçado pela instituição para o período.
- **Alinhamento ao Planejamento Tecnológico:** Conciliar o projeto de abertura dos dados ao planejamento tecnológico do setor de tecnologia da instituição.

4.1 Embasamento Legal

Nesta Seção, estão listados todos os dispositivos legais que subsidiam a abertura dos dados da instituição.

- **Art. 48 da Lei Complementar nº 101, de 4 de maio de 2000:**¹

Adoção de dispositivos de transparência das gestões fiscais por meio eletrônico para acesso público às informações orçamentárias e de prestações de contas.

¹ <http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp101.htm>

- **Decreto Presidencial nº 6.666, de 27 de novembro de 2008:**²

Torna obrigatório, exceto os protegidos por sigilo, o compartilhamento dos dados geoespaciais e seus metadados para todos os órgãos e entidades do Poder Executivo Federal.

- **Lei nº 12.527, de 18 de Novembro de 2011:**³

Criação da Lei de Acesso à Informação (LAI) Regulamenta o direito dos cidadão ao acesso às informações públicas, dispensando a apresentação de qualquer motivo para tal.

- **Instrução Normativa nº4 de 13 de abril de 2012:**⁴

Cria a Infraestrutura Nacional de Dados Abertos (INDA) e estabelece conceitos referentes a: dado, informação, dado público, formato aberto, licença aberta, dados abertos e metadado.

- **Parâmetros estabelecidos na e-PING:**⁵

Institucionaliza os Padrões de Interoperabilidade de Governo Eletrônico (*e-PING*), como a arquitetura de interoperabilidade do governo eletrônico, e os vocabulários e antologias do governo eletrônico *e-VOG*⁶ e *e-MAG*⁷

- **Lei nº 13.460, 26 de junho de 2017:**⁸

Cria uma série de obrigações pro Estado ao passo que empodera o cidadão para participar e para reclamar da qualidade dos serviços públicos. Os dispositivos legais criados são: publicar um quadro geral dos seus serviços (art. 3); detalhamento desses serviços (art. 4); obriga o Estado a prover meios pro cidadão se manifestar sobre os serviços públicos, e as ouvidorias serão o canal oficial de reclamação; cria mecanismos de avaliação continuada pelo cidadão e sugestão de melhorias (formulário *Simplifique*⁹); prevê a criação de conselhos de usuários.

4.2 Motivação Social

É importante citar, além dos dispositivos legais, a motivação social que iniciou todo o processo de abertura dos dados das instituições públicas do país. Diante do cenário de alta demanda dos cidadãos por melhorias dos serviços públicos e combate à corrupção, o *PDA* atua como uma carta à comunidade, e a mensagem não poderia ser mais clara: entendemos a sua demanda, ela é justa e tentaremos ao máximo atender as suas expectativas. Este fato é extremamente relevante, pois a publicação de dados públicos ocorre justamente para atrair o cidadão a acompanhar a atuação do governo e reclamar pelos seus direitos.

² <http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Decreto/D6666.htm>

³ <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>

⁴ <<https://www.governodigital.gov.br/documentos-e-arquivos/3%20-%20IN%2004%2013-04-12.pdf>>

⁵ <https://www.governodigital.gov.br/documentos-e-arquivos/legislacao/Portaria_e-PING_-14_07_2005.pdf>

⁶ *e-VOG* - <<http://vocab.e.gov.br>>

⁷ *e-MAG* - <https://www.governodigital.gov.br/documentos-e-arquivos/legislacao/portaria3_eMAG.pdf>

⁸ <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2017/lei/L13460.htm>

⁹ *Formulário Simplifique* - <<http://www.fazenda.gov.br/ouvidoria/simplifique>>

Além da clara mensagem à população, a publicação de dados públicos fomenta uma atividade científica, no tocante à análise dos dados utilizando ciência. Com os dados das instituições acessíveis pela *web*, a análise científica de dados terá bastante material para se manter ativa, novos métodos de análise dos gastos públicos podem ser desenvolvidos, o que fortalece cada vez mais a transparência governamental. A exemplo disso, pode-se citar o projeto *Operação Serenata de Amor*¹⁰, que é um projeto aberto e independente, mantido por um grupo de cientistas de dados, e que utilizaram técnicas de inteligência artificial e ciência de dados para programar a robô *Rosie*¹¹, que avalia as cotas de atividade parlamentar (*CEAP*) de deputados federais e senadores.

Com o fortalecimento da ciência de dados governamentais, espera-se também o aumento da atividade tecnológica na área, pois, de fato, o governo está se tornando cada vez mais digital. O desenvolvimento de sistemas de informação, na forma de aplicativos *mobile* e sites na *web*, que utilizam essas informações também aumentará. Este fato também contribui para atrair os cidadãos para participarem do governo, também fortalecendo o processo de transparência. Aqui pode-se citar como exemplo a interface amigável para *web* da *Rosie*, o *Jarbas*¹², ou o projeto *Transparência Traduzida* (CAMPOS et al., 2018), citado anteriormente no referencial teórico.

O desenvolvimento de todos esses dispositivos socio-tecnológicos contribui com o aumento da transparência pública, e com a melhoria gradativa na qualidade dos serviços públicos ofertados à população.

4.3 Alinhamento ao Planejamento Estratégico da Instituição

A UFS desenvolve periodicamente um planejamento, denominado *Plano de Desenvolvimento Institucional (PDI)*, que apresenta as diretrizes necessárias a orientar administrativamente e socialmente a instituição.

Tomando com base o *PDI* da UFS, o *PDA* foi desenvolvido consoante com as diretrizes dispostas no item 4 do *PDI*:

- **Melhoria dos indicadores de desempenho acadêmico:** Espera-se que, com a publicação dos dados acadêmicos, os problemas de evasão, reincidente reprovação, dentre outros, possam ser atacados cientificamente com ciência de dados.
- **Infraestrutura física de ensino, pesquisa e extensão, biblioteca, recursos de tecnologia da informação e comunicação:** A implementação do Portal *Dados Abertos UFS* já é uma contribuição no campo tecnológico para esta diretriz.

¹⁰ *Operação Serenata de Amor* - <<https://serenata.ai>>

¹¹ *Rosie da Serenata* - <<https://twitter.com/RosieDaSerenata>>

¹² *Jarbas* - <<https://jarbas.serenata.ai>>

- **Relação com a comunidade acadêmica e com a sociedade:** No que tange à relação com a comunidade em geral, o desenvolvimento do *PDA* e do Portal *Dados Abertos UFS* torna-se um convite à sociedade para acompanhamento das ações institucionais, promovendo o engajamento dos cidadãos, e estreitando as relações da instituição com a comunidade.
- **Gestão organizacional e desenvolvimento de pessoal:** O Portal *Dados Abertos UFS* tem capacidade para armazenar qualquer tipo de *dataset*, e é importante que, futuramente, sejam publicadas informações sobre a gestão e desenvolvimento de pessoas da instituição.
- **Sustentabilidade ambiental e qualidade de vida:** De maneira análoga ao item anterior, é desejável que, futuramente, informações ambientais e de qualidade de vida dos cidadãos universitários sejam publicadas no portal.

Estes dois últimos itens, embora não implementados diretamente, são afetados pela publicação de *datasets* sobre os temas. Qualquer conjunto de dados que possa auxiliar na resolução de problemas da instituição, e da comunidade em geral, é desejável a sua publicação.

De maneira análoga, as mesmas justificativas se aplicam às mesopolíticas elencadas no *PDI* da *UFS*, para as pró-reitorias e diretorias de centro dos campi de São Cristóvão, Aracaju, Laranjeiras, Itabaiana, Lagarto e Nossa Sra. da Glória.

4.4 Alinhamento ao Planejamento Tecnológico da Instituição

O Plano Diretor de Tecnologia de Informação e Comunicação (*PDTIC*) é o documento que estabelece as diretrizes tecnológicas, objetivos e riscos a serem gerenciados para um determinado período.

Consoante com o *PDTIC*, estão as diretrizes tecnológicas a seguir:

- **Prover solução integrada em sistemas de informação**
- **Implantar e manter política de segurança da informação**
- **Definir política de utilização de recursos de Tecnologia de Informação e Comunicação (TIC)**

Como pode-se observar na Seção sobre a Arquitetura *Open Data UFS*, a solução proposta no Capítulo 3, e o Portal que será apresentado no capítulo seguinte, obedecem às diretrizes requeridas pelo planejamento tecnológico da instituição, e portanto justificam-se como projeto tecnológico viável para a *UFS*.

4.5 Desenvolvimento do PDA e Execução

O desenvolvimento do *PDA* utilizou a arquitetura desenvolvida neste trabalho. Também foram considerados como referência as orientações do *Manual Para Elaboração de Plano de Dados Abertos*, desenvolvido pelo Ministério do Planejamento, Orçamento e Gestão (MPOG, 2015). Para a confecção do documento *PDA* também foram consideradas as diretrizes legais, sociais e de estratégia da instituição e tecnológicas, conforme apresentado nas seções anteriores.

A partir da arquitetura de dados abertos proposta no capítulo 3, foram possíveis as seguintes ações que facilitaram a elaboração do *PDA*:

- A identificação das partes interessadas e seus respectivos interesses, definidos no capítulo anterior, seções 3.2 e 3.3, permitiram conhecer melhor o público-alvo do portal e, dessa forma, selecionar os *datasets* e seus atributos com foco nos interesses (*concerns*) dos visitantes do portal.
- O ponto de vista do negócio descrito na arquitetura de dados abertos apresentada no capítulo anterior, seção 3.5.1, permitiu aliar os interesses gerais com os interesses tecnológicos da instituição, a fim de justificar a implantação de uma política de transparência e publicação de dados.
- Os modelos sugeridos na visão de dados do capítulo anterior, seção 3.5, subseção 3.5.2, deram origem ao modelo lógico de dados, incluído do *PDA* para orientar os usuários do portal quanto aos metadados das informações disponibilizadas. Na visão de dados citada, também são estabelecidos critérios para a seleção das informações a serem extraídas e publicadas. Por fim, nesta visão também são definidos filtros que impedem a divulgação das informações, com a finalidade de preservar a integridade institucional e a individual dos cidadãos universitários.

Seria impossível catalogar e publicar todos os dados disponíveis na instituição, pois alguns estão de fácil acesso, em bases de dados com regras de negócio bem conhecidas pela equipe de sistemas. Porém, outras informações estão em bases com pouco acesso e manutenção da equipe, e se o acesso a ela não estiver bem documentado, será necessário um processo de engenharia reversa para resgate da informação. Outras informações podem sequer estar ao alcance da equipe de tecnologia do *NTI*, e podem estar em algum repositório particular de setores da instituição. E no pior caso podem existir informações que nem estão digitalizadas, e portanto estão em formato físico arquivadas em algum setor da instituição.

Diante da situação, foi adotada a metodologia de publicação iterativa: os dados serão abertos em pequenos grupos a cada iteração, atendendo a ordem de prioridade definida na arquitetura de dados abertos, apresentada no capítulo anterior, seção 3.5, e subseção 3.5.2, retomadas a seguir:

- Necessidade de cumprimento da legislação (peso 4).
- Consistência de informação (peso 3).
- Relevância para a comunidade (peso 2).
- Facilidade de acesso à informação para consulta (peso 1).

Além da ordem de prioridade e dos pesos, no capítulo anterior, seção 3.5, e subseção 3.5.2, também foram definidos os filtros de impedimento da publicação:

Também é importante definir filtros de impedimento da publicação das informações como:

- Não divulgar dados pessoais dos indivíduos cadastrados nas bases da instituição.
- Não divulgar dados sigilosos da instituição.

4.6 Aprovação do PDA

Após a confecção do documento, este foi submetido para avaliação pela Comissão de Tecnologia da Informação (CTIn). De acordo com a Portaria 0119 GR da UFS¹³ que estabelece a composição e o regulamento da Comissão de Tecnologia da Informação CTIn, ela é composta por:

- **Pró-Reitor de Planejamento**
- **Pró-Reitor de Administração**
- **Pró-Reitor de Assuntos Estudantis**
- **Pró-Reitor de Pós-graduação e Pesquisa**
- **Pró-Reitor de Extensão e Assuntos Comunitários**
- **Pró-Reitor de Graduação**
- **Diretor do Núcleo de Tecnologia da Informação (NTI)**
- **Coordenador de Sistemas do NTI**
- **Coordenador de Redes do NTI**
- **Coordenador do NPQ (Núcleo de Promoção de Qualidade)**

¹³ Portaria 0119 GR da UFS - <http://portarias.ufs.br/uploads/content_attach/path/15104/0119_Comiss_o_de_Tecnologia_da_Informa__o_CTIn_.pdf>

- **Responsável Jurídico**
- **Secretário**
- **Assessor da Reitoria para assuntos de TI**

No dia 27 de novembro de 2018 o Plano de Dados Abertos *PDA* da *UFS* foi submetido à avaliação pela distinta comissão gestora de tecnologia da instituição, o *CTIn* e logrou êxito em aprovação. Em seguida, o documento foi remetido à *Controladoria Geral da União (CGU)*, que recomendou que fossem implementados os seguintes itens como melhoria:

- Inventário de bases de dados do órgão independente de serem submetidas a abertura ou não.
- Das bases de dados que serão submetidas a abertura, elaborar um cronograma.
- Mecanismos de participação social na priorização da abertura de bases.
- Cronograma de mecanismos para promoção e fomento.

Os itens acima recomendados pela *CGU* estão atualmente em processo de esclarecimento e desenvolvimento junto a *AUDINT*, setor de auditoria e controle interno da instituição, e *CTIn*.

4.7 Desafios Superados e Sugestões para o Próximo PDA

Durante o desenvolvimento do *PDA*, a instituição recebeu uma notificação da Controladoria Geral da União (*CGU*) cobrando posicionamento acerca do desenvolvimento do documento e implantação da política de Dados Abertos da instituição.

Até então, o desenvolvimento do *PDA* e do Portal *Dados Abertos UFS* não era nem de conhecimento dos setores de transparência responsáveis, e se não fosse pelo desenvolvimento deste trabalho, a instituição provavelmente seria autuada pelo órgão de controle. Portanto a comunicação entre as partes interessadas no projeto deve ser grande com o objetivo de superar falhas de entendimento, e não causarem descumprimento de prazos. Além das falhas de comunicação, a identificação dos *datasets* também foi um problema durante o projeto, pois algumas funcionalidades dos sistemas *SIG* nem estavam sendo operadas pelos usuários responsáveis, dessa forma, desatualizando as informações necessárias para extração. Outro grande problema, foi encontrar os documentos oficiais de planejamento estratégico da instituição e planejamento tecnológico da instituição. Durante a pesquisa, ocorreu retrabalho em algumas fases pela descoberta de documentos mais atuais. Portanto, foi solicitado que a instituição mantenha todos esses documentos importantes disponibilizados numa área para download em seu portal.

A seguir, será apresentada a lista de *datasets*, e desde já fica como sugestão para o próximo documento a inserção da lista de *datasets* disponibilizados no portal. Outros *datasets* estavam cogitados para entrar nessa lista, porém as tabelas necessárias para extração das informações não estavam sendo alimentadas pelos usuários da instituição nas bases de dados do *SIGAA* e *SIPAC*.

- **Datasets homologados para publicação:** Empenhos; Departamentos; Cursos; Turmas; Docentes; Discentes; Restaurantes; Cardápio de Restaurantes; Telefones.
- **Datasets não-homologados para publicação:** Liquidação de Empenho; Ordens Bancárias; Contratos; Convênios; Diárias; Passagens; Atividades de Extensão; Projetos de Pesquisa.

A justificativa dada pelos usuários é que eles utilizam outros sistemas para gerenciar essas informações e a alimentação destas no *SIGAA* e *SIPAC* demanda retrabalho. De fato, a realimentação de informações em dois sistemas distintos consome tempo, porém a não alimentação dessas informações no sistemas adequados impede a sua exportação via *API Sistemas*. Para corrigir essa falha, a sugestão é que sejam criadas rotinas que automatizem a exportação das informações para as bases do *SIGAA* e *SIPAC*.

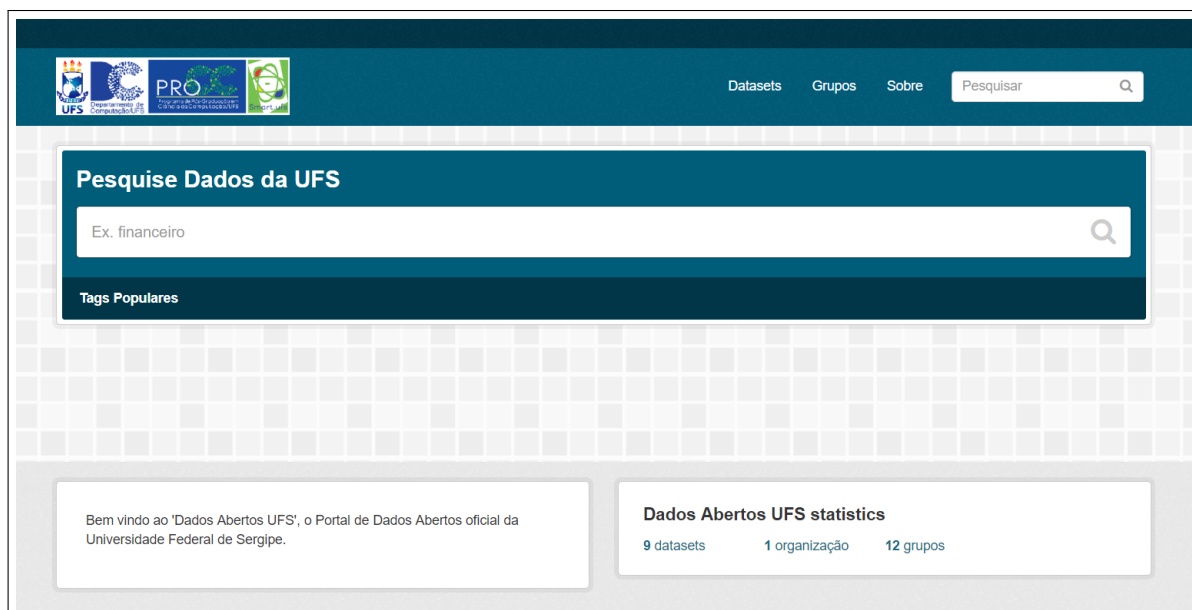
4.8 Apresentação do Portal de Dados Abertos da UFS

Ao acessar o portal *Dados Abertos UFS*, através do endereço <http://dados.ufs.br> ou <http://opendata.smart.ufs.br>, o visitante se depara com uma interface minimalista, propositalmente selecionada, para conduzi-lo a utilizar a barra de pesquisa, ou o menu superior direito que contém três funções: *Datasets*, *Grupos* e *Sobre*. Esta interface foi escolhida com o objetivo de reduzir a poluição visual que costuma confundir o usuário, e simplificar os passos executados para acessar as informações desejadas. A Figura 15 mostra a página principal do portal.

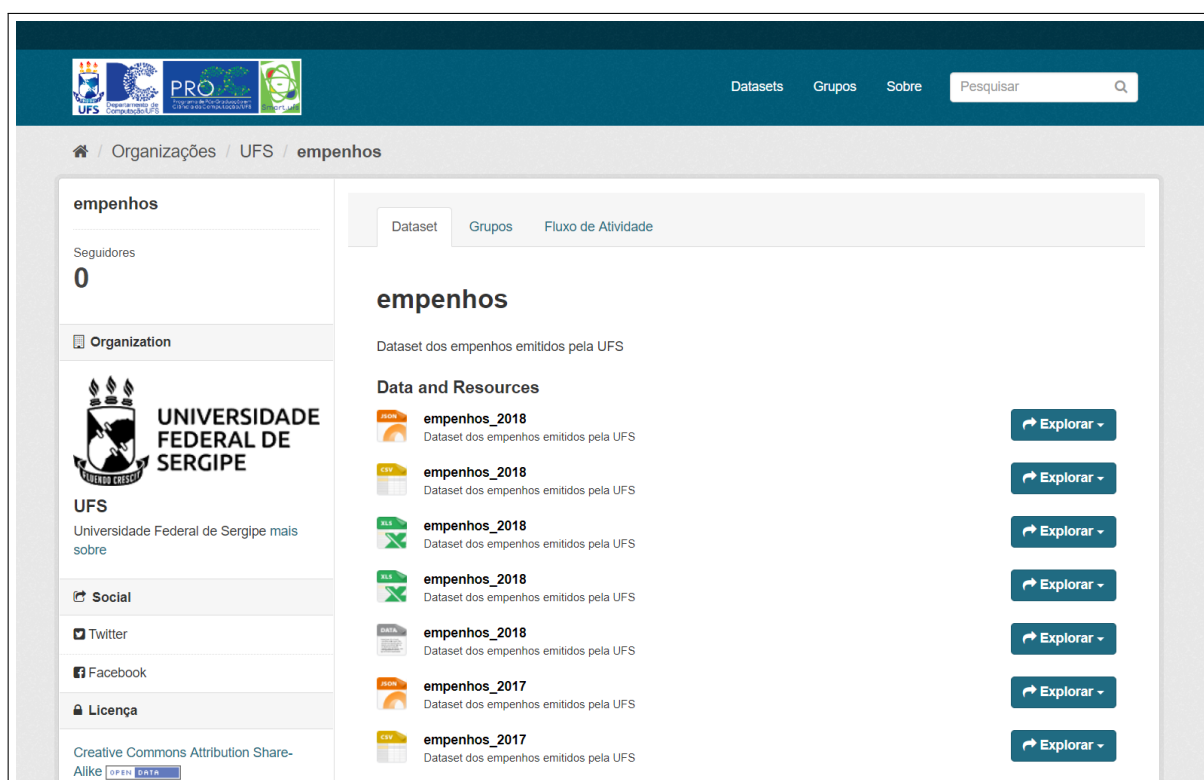
No menu *Datasets*, o usuário encontra a lista de *datasets* disponíveis e ao escolher um deles, o portal o redireciona para uma lista de arquivos publicados, em vários formatos. Ao clicar em um arquivo o usuário é redirecionado para a tela de pré-visualização dos dados e *download* do arquivo. O usuário também pode utilizar o botão *Explorar* para realizar estas ações. As Figuras 16 e 17 ilustram a ação. Ainda na tela de pré-visualização do arquivo, podemos perceber que o *CKAN* possui outras funcionalidades interessantes, como mapa, gráfico de funções e pesquisa de informações.

A função de mapas pode ser utilizada quando o *dataset* possui coordenadas e outras informações geográficas. O usuário informa ao *CKAN* quais as colunas de latitude e longitude, e então o *software* exibe no mapa o ponto resultante.

A função de gráfico, análoga à de mapa, funciona quando o usuário seleciona o tipo de gráfico que deseja gerar, e as colunas que farão parte do gráfico da função.

Figura 15 – Visão geral do portal *Dados Abertos UFS*

Fonte - Autores (2018)

Figura 16 – Tela de lista de arquivos de um *dataset*

Fonte - Autores (2018)

Por fim, a tela de pré-visualização também possui outra função interessante, que é a de buscar informação num *dataset*. Esta é uma mera comodidade para evitar que o usuário realize o *download* do *dataset* apenas para buscar informações.

Figura 17 – Tela de pré-visualização e download de um *dataset*

empenhos_2018 Download

URL: <http://opendata.smart.ufs.br/dataset/80a82c28-15fb-496a-9ce9-c6638b797622/resource/f97dc7a8-9bd7-428f-b7d4-ccedbd711feb/download/empenhos.xls>

Dataset dos empenhos emitidos pela UFS

Data Explorer

Fullscreen Embed

Grid Graph Map 838 records « 1 – 100 »

Search data ... Go » Filters

column_1	ano	anulado	credor	data	id_ug	nome_ug	nro_em...	observac...	saldo	sigla_uf
0	2018	0	FUNDAÇÃO UNIVERSID...	Apr 23, 2018 11:21:...	605	FUNDA...	210	DIARIAS ...	30000	UFS
1	2018	0	PACIFIC ORGANIZADOR...	Apr 23, 2018 10:57:...	605	FUNDA...	800246	CONTRA...	17000	UFS
2	2018	0	RC RAMOS COMERCIO ...	Apr 24, 2018 5:05:0...	605	FUNDA...	800261	AQUISIC...	18360	UFS
3	2018	0	JOAO BOSCO GOMES	Apr 5, 2018 5:03:00...	605	FUNDA...	800216	EMP SU...	0	UFS
4	2018	0	TICKET SOLUCOES HD...	Apr 24, 2018 12:00:...	605	FUNDA...	800257	ADESÃO...	582057.12	UFS
5	2018	0	VIA RETA COMERCIO, S...	Apr 25, 2018 11:37:...	605	FUNDA...	800263	EMPENH...	70274.37	UFS
6	2018	0	FUNDAÇÃO UNIVERSID...	Apr 27, 2018 2:09:0...	605	FUNDA...	230	PAGAME...	1202.58	UFS
7	2018	0	FUNDAÇÃO UNIVERSID...	Apr 27, 2018 4:06:0...	605	FUNDA...	232	DIARIAS ...	51857.74	UFS
8	2018	0	FUNDAÇÃO UNIVERSID...	Apr 27, 2018 4:14:0...	605	FUNDA...	234	AUXILIO ...	250000	UFS
9	2018	0	FUNDAÇÃO UNIVERSID...	Apr 27, 2018 4:19:0...	605	FUNDA...	235	AJUDA D...	300000	UFS
10	2018	0	MOURA EMPREENDIME...	Apr 27, 2018 3:26:0...	605	FUNDA...	800272	ADITIVO ...	7008	UFS
11	2018	0	BANCO DO BRASIL SA	May 2, 2018 9:29:0...	605	FUNDA...	800273	PASSAG...	200000	UFS
12	2018	0	UNIAO BRASILEIRA DE ...	May 2, 2018 9:55:0...	605	FUNDA...	236	REEMBO...	1400	UFS
13	2018	0	FEDERACAO ATLETICA ...	May 4, 2018 10:03:...	605	FUNDA...	237	TAXA DE...	7994	UFS

Fonte - Autores (2018)

4.9 Considerações Finais do Capítulo

Neste capítulo foram apresentadas as diretrizes que nortearam a elaboração do Plano de Dados Abertos, o *PDA*, da *UFS*. Foi apresentado que o documento utiliza-se de dispositivos legais, sociais, e diretrizes de interesse institucional no âmbito administrativo e tecnológico. Por fim, foi mostrado que apesar das dificuldades encontradas durante o processo de elaboração do plano, estas foram superadas e o documento logrou êxito em aprovação pelo *CTIn*. O *PDA* desenvolvido encontra-se no Apêndice A desta dissertação.

Por fim, foi apresentada, neste capítulo, a construção do portal *Dados Abertos UFS*, como também as tecnologias envolvidas para o pleno funcionamento da ferramenta *CKAN* que serve de base para personalização do portal. Além disso, algumas telas do resultado da personalização foram apresentadas, mostrando algumas funcionalidades importantes.

É importante concluir também que o portal *Dados Abertos UFS*, de acordo com o modelo *5-Star* (HAUSENBLAS; KIM, 2012), eleva o nível de abertura da instituição do primeiro nível para o terceiro (três estrelas), que é o nível de publicação de dados nos formatos abertos de datasets.

5

Conclusão

Neste trabalho foram estudados os conceitos de dados abertos no capítulo 2, e demonstrada a importância de aprovar um plano de dados abertos para subsidiar o processo de abertura dos dados de uma instituição. Com base na arquitetura de dados abertos proposta no capítulo 3 foi possível desenvolver o *Plano de Dados Abertos (PDA)* da UFS, a solução *Open Data Link* de extração dos dados dos sistemas da instituição e publicação no portal *Dados Abertos UFS*. Também com base na arquitetura implementada, para reunir e catalogar os datasets da instituição, foi desenvolvido o portal *Dados Abertos UFS* com base no *software CKAN*. Logo a hipótese 1 definida no capítulo 1, seção 1.4, é verdadeira.

Com relação aos desafios encontrados, as dificuldades em desenvolver o *PDA* foram devidamente demonstradas no capítulo 4. As dificuldades de implementação da arquitetura, do portal e da solução *Open Data Link* foram registradas no capítulo 3.

Os resultados da pesquisa mostram que a criação e aprovação do *PDA* de fato orientou e facilitou o processo de abertura dos dados. Logo a hipótese 2 do capítulo 1, seção 1.4, é verdadeira. A devida aprovação do *PDA* pela comissão de tecnologia da instituição, o *CTIn* indica que houve êxito na aceitação inicial do documento pela instituição, muito embora não tenha sido possível medir a aceitação pelos cidadãos universitários. O *PDA* também passou pela aceitação da *Controladoria Geral da União (CGU)*, que sugeriu algumas mudanças para melhoria do documento, que estão atualmente em implementação. Além disso, na escala 5 *Estrelas* (HAUSENBLAS; KIM, 2012) de abertura de dados, o portal *Dados Abertos UFS* eleva a instituição para o terceiro nível de transparência.

5.1 Principais Contribuições

As principais contribuições deste trabalho estão dispostas a seguir, mostrando que o objetivo geral, assim como os objetivos específicos foram alcançados:

- Desenvolvimento da arquitetura de dados abertos proposta com base na norma *ISO/IEC/J-EEE 42010:2011* que auxiliou no processo de elaboração do *PDA* da *UFS*, e contribuiu com o processo de transparência da instituição.
- Desenvolvimento do *Plano de Dados Abertos (PDA)* da *UFS* a partir do estudo dos conceitos de dados abertos, superação dos desafios de desenvolver uma política de dados abertos, e aplicação da arquitetura de dados abertos proposta. O *PDA*, em conformidade com o disposto na lei, e alinhado às diretrizes institucionais, tecnológicas e sociais, será o documento oficial que dispõe sobre a política de publicação de dados da instituição de agora em diante.
- Desenvolvimento do portal *Dados Abertos UFS*, que será o portal oficial de fornecimento de informações no formato de *datasets*.
- Desenvolvimento de bibliotecas reutilizáveis para extração (*lib_apisistemas*) e publicação (*lib_ckanUFS*) dos dados da *UFS*, e que utilizam interfaces *RESTful*.
- Desenvolvimento de consultas para a interface *RESTful* da *API Sistemas*.

5.2 Limitações da Pesquisa

A pesquisa limitou-se aqui a extrair os dados que estavam nas bases de dados dos sistemas relacionais da instituição. Porém, conforme relatado no capítulo 4, os dados sobre liquidações de empenho e pagamento de ordens bancárias, por exemplo, não estão sendo alimentados das bases alcançadas pela *API Sistemas*. Dessa forma esses conjuntos de dados, e outros que estão em outras bases fora da *API Sistemas* foram excluídos do processo.

O portal *Dados Abertos UFS*, de acordo com o modelo *5-star* de abertura dos dados, alcança apenas o nível de 3 estrelas, pois limita-se a publicar os dados em formatos abertos mas não insere as informações num contexto e as define através de *URIs* para definir as entidades.

5.3 Trabalhos futuros

Para trabalhos futuros, a pesquisa pode ir além dos dados alcançados pela *API Sistemas*, e focar em uma maneira de integrar outras bases na *API* de consulta existente ou numa *API* integradora de bases.

É possível também evoluir a arquitetura implementada com o objetivo de contemplar o grau máximo de abertura definido na classificação *5 Estrelas* (HAUSENBLAS; KIM, 2012). Isso demandará uma revisão completa da arquitetura, pois serão necessárias também revisões do negócio, nova catalogação dos dados, uma nova identificação de tecnologias dentro da área de *web* semântica. Com essa revisão, novas partes interessadas e respectivos interesses, visões,

e requisitos também deverão surgir, demandando por consequência uma revisão do *PDA* e refatoração das bibliotecas de extração e carga no portal. Todo este processo configura material suficiente para uma nova pesquisa.

Uma revisão sistemática também pode ser realizada para captar o estado da arte na área, pois novas arquiteturas de dados abertos e frameworks de implantação podem ter surgido durante a elaboração desta pesquisa.

Referências

ALBINO, V.; BERARDI, U.; DANGELICO, R. M. Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology*, v. 22, n. 1, p. 3–21, 2015. Citado na página 18.

ANDERSEN, T. B. E-government as an anti-corruption strategy. *Information Economics and Policy*, v. 21, n. 3, p. 201 – 210, 2009. ISSN 0167-6245. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0167624509000110>. Citado na página 10.

ATTARD, J. et al. A systematic review of open government data initiatives. *Government Information Quarterly*, v. 32, n. 4, p. 399 – 418, 2015. ISSN 0740-624X. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0740624X1500091X>. Citado 4 vezes nas páginas 11, 15, 18 e 19.

ATZORI, L.; IERA, A.; MORABITO, G. The internet of things: A survey. *Computer Networks*, v. 54, n. 15, p. 2787 – 2805, 2010. ISSN 1389-1286. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1389128610001568>. Citado na página 19.

BAKICI, T.; ALMIRALL, E.; WAREHAM, J. A smart city initiative: the case of barcelona. *Journal of the Knowledge Economy*, v. 4, n. 2, p. 135–148, Jun 2013. ISSN 1868-7873. Disponível em: <https://doi.org/10.1007/s13132-012-0084-9>. Citado na página 18.

BARBALHO, F. A. *Emergência de um campo de ação estratégica: o caso de política pública sobre dados abertos*. 253 p. Tese (Doutorado em Administração) — Programa de Pós-Graduação em Administração, Faculdade de Economia, Administração e Contabilidade, Universidade de Brasília, Brasília, 2014. Citado na página 18.

BERTOT, J. C. et al. Big data, open government and e-government: Issues, policies and recommendations. *Information Polity*, IOS Press, v. 19, n. 1, 2, p. 5–16, 2014. Citado na página 20.

CAMPOS, O. S. F. et al. *Transparência Traduzida*. 2018. Disponível em: <http://www.transparenciatraduzida.com.br>. Citado 3 vezes nas páginas 19, 20 e 54.

CHEN, T. M. Smart grids, smart cities need better networks [editor's note]. *IEEE Network*, v. 24, n. 2, p. 2–3, March 2010. ISSN 0890-8044. Citado na página 18.

Copenhagen Technology and Environment Administration. *Copenhagen Data*. 2013. Disponível em: <https://data.kk.dk>. Citado na página 23.

CORREA, A. S. et al. An add-oriented software architecture for structuring information to open government data. *Journal of Computer Science*, v. 14, p. 673–679, 01 2018. Citado na página 15.

DAWES, S. S.; HELBIG, N. Information strategies for open government: Challenges and prospects for deriving public value from government transparency. In: _____. *Electronic Government: 9th IFIP WG 8.5 International Conference, EGOV 2010, Lausanne, Switzerland, August 29 - September 2, 2010. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 50–60. ISBN 978-3-642-14799-9. Disponível em: https://doi.org/10.1007/978-3-642-14799-9_5. Citado na página 18.

FIELDING, R. T. *Architectural styles and the design of network-based software architectures*. [S.l.]: University of California, Irvine Irvine, USA, 2000. v. 7. Citado 2 vezes nas páginas 30 e 31.

GUEDES, G. *UML 2 - Uma Abordagem Prática*. [S.l.]: Novatec Editora, 2018. ISBN 9788575226445. Citado na página 28.

HAUSENBLAS, M.; KIM, J. G. B. *5 Estrelas dos Dados Abertos*. 2012. Disponível em: <<https://5stardata.info>>. Citado 5 vezes nas páginas 21, 37, 61, 62 e 63.

HOFMEISTER, C.; NORD, R.; SONI, D. *Applied Software Architecture*. [S.l.]: Addison Wesley, 2000. Citado na página 33.

International Organization Of Standardization. *ISO/IEC 10746-1 - Information technology open distributed processing reference model: overview*. [S.l.], 1998. Citado na página 37.

International Organization Of Standardization. *ISO/IEC/IEEE 42010:2011 - Systems and software engineering – Architecture description*. [S.l.], 2011. v. 2011, 1–46 p. Disponível em: <[http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6129467&contentType=Standards&sortType=asc_p_Sequence&filter=AND\(p_Publication_Number:6129465\)>](http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6129467&contentType=Standards&sortType=asc_p_Sequence&filter=AND(p_Publication_Number:6129465)>)>. Citado 10 vezes nas páginas 25, 26, 28, 32, 33, 35, 38, 39, 40 e 41.

KONG, X. Y. *Arquitetura para publicação de dados abertos em sistemas de governo: O exemplo do SIASG*. 90 p. Dissertação (Mestrado em Engenharia de Sistemas e Computação) — Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2015. Citado 2 vezes nas páginas 14 e 15.

KOURTIT, K.; NIJKAMP, P. Smart cities in the innovation age. *Innovation: The European Journal of Social Science Research*, v. 25, n. 2, p. 93–95, 2012. Citado na página 18.

KRUCHTEN, P. The 4+1 view model of architecture. *IEEE Software*, v. 12, p. 45–50, 11 1995. Citado na página 33.

Lucid Chart. *Component Diagram Tutorial*. 2019. Disponível em: <<https://www.lucidchart.com/pages/uml-component-diagram>>. Citado na página 29.

MPOG, M. *Manual para Elaboração de Plano de Dados Abertos*. 2015. Disponível em: <http://www.planejamento.gov.br/secretarias/upload/arquivo/governo-aberto/copy_of_manual_elaboracao_plano_dados_abertos.pdf/@@download/file/manual_elaboracao_plano_dados_abertos.pdf>. Citado 2 vezes nas páginas 14 e 56.

MURILLO, M. J. Evaluating the role of online data availability: The case of economic and institutional transparency in sixteen latin american nations. *International Political Science Review*, v. 36, n. 1, p. 42–59, 2015. Citado na página 18.

MUSSKOPF, I.; CUDUCOS, E.; CABRAL, F. *Operação Serenata de Amor*. 2016. Disponível em: <<https://serenata.ai>>. Citado na página 10.

OKI, O. *O que são Dados Abertos?* 2011. Disponível em: <http://opendatahandbook.org/guide/pt_BR/what-is-open-data/>. Citado na página 20.

Open Knowledge International. *About CKAN*. 2006. Disponível em: <<https://ckan.org/about/>>. Citado na página 24.

OWIECIŃSKA, K.; LEGIERSKI, J. Open data collection using mobile phones based on ckan platform. In: . [s.n.], 2015. p. 1191–1196. ISBN 9788360810651. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84958063979&doi=10.15439%2f2015F128&partnerID=40&md5=c68c0a12e242565b5e9cca3bbd194590>>. Citado na página 22.

POSSAMAI, A. J. *Dados abertos no Governo Federal brasileiro: desafios de transparência e interoperabilidade*. 313 p. Tese (Doutorado em Ciência Política) — Programa de Pós-Graduação em Ciência Política, Instituto de Filosofia e Ciências Humanas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016. Citado 2 vezes nas páginas 14 e 17.

ROSS, J. W.; WEILL, P.; ROBERTSON, D. C. *Arquitetura de TI como estratégia empresarial*. [S.l.]: M. Books, 2008. Citado na página 32.

SEPLAG-AL, S. *Alagoas em Dados e Informações*. 2013. Disponível em: <<http://dados.al.gov.br>>. Citado na página 23.

SILVA, E. L.; MENEZES, E. M. *Metodologia da Pesquisa e Elaboração de Dissertação*. Santa Catarina: Universidade Federal de Santa Catarina, 2001. Citado na página 13.

SOCRATA. *Open Data Field Guide*. 2015. Disponível em: <<https://socrata.com/open-data-field-guide>>. Citado na página 14.

SOMMERVILLE, I. *Software Engineering*. 9. ed. Boston, MA, USA: Addison-Wesley, 1992. Citado na página 32.

The Open Group. *TOGAF Version 9.2*. [S.l.], 2018. Citado 5 vezes nas páginas 33, 38, 39, 40 e 41.

TOMAS, G. H. R. P. *Uma arquitetura para cidades inteligentes baseada na internet das coisas*. 109 p. Dissertação (Mestrado em Ciência da Computação) — Programa de Pós-Graduação em Ciência da Computação, Centro de Informática, Universidade Federal de Pernambuco, Recife, 2014. Citado na página 19.

Transparency International. *Corruption Perception Index 2018*. 2018. Disponível em: <<https://www.transparency.org/country/BRA>>. Citado na página 18.

United Kingdom Ministry of Defence. *MOD Architecture Framework*. [S.l.], 2012. Disponível em: <<https://www.gov.uk/guidance/mod-architecture-framework>>. Citado na página 33.

VASSILIADIS, P. A survey of extract–transform–load technology. *International Journal of Data Warehousing Mining*, IGI Global, v. 5, n. 3, p. 1–27, 2009. Citado na página 48.

Apêndices

APÊNDICE A – Plano de Dados Abertos (PDA) da UFS

Este PDA foi desenvolvido como parte do projeto apresentado, com o intuito de definir um planejamento para a sistemática abertura dos dados da Universidade Federal de Sergipe. Além disso, o documento também serve de referência para os envolvidos no processo de abertura dos dados da instituição.



PLANO DE DADOS ABERTOS DA UFS

2018 - 2020

Histórico de Revisões

[illegible]

Sumário

1. Apresentação	3
2. Introdução	4
2.1. Dispositivos Legais	4
2.2. Diretrizes Institucionais	5
2.2.1. Plano de Desenvolvimento Institucional	5
2.2.2. Plano Diretor de Tec. da Informação (PDTIC)	6
2.3. Objetivo geral	6
2.4. Objetivos específicos	6
2.5. Projeto SmartUFS	7
3. Critérios de Abertura	8
3.1. Facilidade de Acesso à informação	8
3.2. Necessidade de cumprimento dos dispositivos legais	8
3.3. Relevância da publicação para a comunidade	8
3.4. Consistência da informação	8
4. Estratégias e Premissas para Abertura	9
4.1. Estratégias	9
4.2. Premissas	9
5. Catálogo de Datasets Publicados	11
6. Catalogação no Portal Brasileiro de Dados Abertos	14
7. Sustentação	15
a. Governança	15
8. Monitoramento e controle	16
a. Melhoria da qualidade dos dados	16
b. Comunicação e participação social	16
9. Plano de ação	17
Glossário	19

1. Apresentação

Este documento refere-se ao plano de dados abertos da Universidade Federal de Sergipe (UFS). O Plano de Dados Abertos (PDA) é o documento que descreve os conjuntos de dados (datasets) relativos à UFS que serão disponibilizados ao público, e determina outras informações importantes como elenca os dispositivos legais e objetivos estratégicos que dão subsídio à publicação, e a periodicidade de atualização das informações.

O PDA deve ser aprovado pela Comissão de Tecnologia da Informação (CTIn) da UFS e revisado a cada dois anos, ou em momento oportuno, se antes ocorrer. A partir da sua data de publicação, todas as unidades da instituição devem colaborar, conforme demanda, para que a instituição cumpra com o seu dever de prover as informações institucionais no mais alto nível de qualidade possível.

Os dados devem ser publicados na maior quantidade de formatos possíveis e necessária para promover a transparência, e democratização no uso de ferramentas de leitura. O objetivo dessa meta é atingir o maior público possível, interessado nos dados da instituição. A UFS também se compromete a não publicar informações que não sejam públicas ou que sejam de caráter restrito.

Com a publicação dos dados institucionais, a UFS espera que o público em geral faça bom uso deles, seja mais fiscalizador e participativo na sua gestão. Além do foco na transparência, a instituição também pretende atrair desenvolvedores, empresários e outros órgãos da administração pública, a fim de utilizarem estes dados e evidenciar situações de interesse coletivo.

2. Introdução

A UFS institui, por meio deste documento, o plano inicial de abertura dos seus dados, e descreve os procedimentos e dados sob sua responsabilidade e leva em consideração os dispositivos legais e diretrizes institucionais abaixo:

2.1. Dispositivos Legais

- O disposto no art. 48 da Lei Complementar nº 101, de 4 de maio de 2000¹, que determina ao Poder Público a adoção de instrumentos de transparência na gestão fiscal em meios eletrônicos de acesso público às informações orçamentárias e prestações de contas.
- O Decreto Presidencial nº 6.666, de 27 de novembro de 2008², que instituiu a criação da Infraestrutura Nacional de Dados Espaciais – INDE e determina que o compartilhamento e disseminação dos dados geoespaciais e seus metadados é obrigatório para todos os órgãos e entidades do Poder Executivo Federal, salvo os protegidos por sigilo.
- O Decreto s/nº de 15 de setembro de 2011³, que institui o Plano de Ação Nacional sobre Governo Aberto, o qual estabelece o compromisso do governo de implantar a Infraestrutura Nacional de Dados Abertos (INDA).
- A Instrução Normativa nº4 de 13 de abril de 2012⁴, que cria a INDA e estabelece conceitos referentes a: dado, informação, dado público, formato aberto, licença aberta, dados abertos e metadado.
- O Plano de Ação da INDA⁵, que institui a necessidade dos órgãos de instituírem seus respectivos Planos de Abertura de Dados com vistas a uma Política Nacional de Dados Abertos e institui os elementos mínimos do documento, bem como orienta que a abertura de dados deve observar a relevância para o cidadão.
- A Parceria para Governo Aberto (Open Government Partnership – OGP), celebrada em setembro de 2011 entre o Brasil e sete outros países (mencionada acima), que pactuou novos compromissos a partir de 2013 (2º Plano de Ação⁶).

¹ http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp101.htm

² http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Decreto/D6666.htm

³ http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Dsn/Dsn13117.htm

⁴

<http://www.governoeletronico.gov.br/biblioteca/arquivos/instrucao-normativa-da-infraestrutura-nacional-de-dados-abertos-2013-inda/download>

⁵

<http://wiki.gtinda.ibge.gov.br/GetFile.aspx?File=%2fGT1-Gest%c3%a3o%20e%20Normativo%2fPlano%20de%20A%c3%a7%c3%a3o%20-%20INDA%2005.02.2013.odt>

⁶

http://www.cgu.gov.br/PrevencaodaCorrupcao/CompromissosInternacionais/GovernoAberto/documentos/arquivos/ogp-brasil-plano_acao2.pdf

- A criação e entrada em vigor da Lei nº 12.527, de 18 de Novembro de 2011⁷, a chamada Lei de Acesso à Informação – LAI.
- Os parâmetros estabelecidos na e-PING⁸ - arquitetura de interoperabilidade do governo eletrônico, e os vocabulários e ontologias de Governo Eletrônico e-VoG⁹ e e-MAG - Modelo de Acessibilidade de Governo Eletrônico¹⁰ (instituído pela Portaria nº 03, de 07 de Maio de 2007).
- Os parâmetros atuais ou que venham a ser estabelecidos no âmbito de Planejamento Estratégico Institucional¹¹ ou setorial, bem como os relacionados às áreas de tecnologia da informação (PETI e PDTI¹²), sob orientação da Estratégia Geral de Tecnologia da Informação – EGTI¹³.

2.2. Diretrizes Institucionais

A UFS tem sua sede central na Cidade Universitária Prof. José Aloísio de Campos, no bairro Rosa Elze, município de São Cristóvão, do estado de Sergipe. Foi criada e mantida pela União sob a forma de fundação, nos termos do Decreto – Lei nº 269 de 28 de fevereiro de 1967, integrada ao Sistema Federal de Ensino Superior Brasileiro com a incorporação dos cursos superiores do Estado à época. (PDTIC UFS, 2017)

Atualmente, o Núcleo de Tecnologia da Informação (NTI), como unidade responsável pela implantação e manutenção da tecnologia, na UFS tem a responsabilidade de alinhar a estratégia da instituição com as tecnologias que surgem, sempre muito dinâmicas, provendo as ferramentas necessárias para que a UFS cumpra a sua missão para com a sociedade.

2.2.1. Plano de Desenvolvimento Institucional

O Plano de Desenvolvimento Institucional é o documento que reafirma a missão, visão, políticas acadêmicas e administrativas fundamentadas pela realidade institucional e estabelece metas estratégicas para o período compreendido entre os exercícios acadêmicos de 2016 a 2020. (PDI UFS, 2016)

O PDA, enquanto diretriz institucional encontra-se alinhado às seguintes dimensões postas no item 4 do PDI da UFS, elencadas abaixo:

⁷ http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm

⁸ www.governoeletronico.gov.br e

<http://www.governoeletronico.gov.br/o-gov.br/legislacao/portaria-no-05-de-14-de-julho-de-2005>

⁹ <http://vocab.e.gov.br>

¹⁰ <http://www.governoeletronico.gov.br/acoes-e-projetos/e-MAG>

¹¹ http://www.planejamento.gov.br/secretarias/upload/Arquivos/publicacao/planejamento_estrategico/130314_planejamento_estrategico.pdf

¹² PETI – Planejamento Estratégico da Tecnologia da Informação -

http://www.planejamento.gov.br/secretarias/upload/Arquivos/publicacao/20130621_PETI_-_Revisao_Final_DTISE.pdf; PDTI – Plano Diretor de Tecnologia da Informação -

<http://www.planejamento.gov.br/secretarias/upload/Arquivos/PDTI.pdf>

¹³ http://www.sisp.gov.br/egti/wiki/download/file/EGTI_2013-2015_v1_1.pdf

- Melhoria dos indicadores de desempenho acadêmico.
- Infraestrutura física de ensino, pesquisa e extensão, biblioteca, recursos de tecnologia da informação e comunicação.
- Relação com a comunidade acadêmica e com a sociedade.
- Gestão organizacional e desenvolvimento de pessoal.
- Sustentabilidade ambiental e qualidade de vida.

O PDA também se encontra em sintonia com as mesopolíticas elencadas no PDI da UFS, para as pró-reitorias e diretorias de centro dos campi de São Cristóvão, Aracaju, Laranjeiras, Itabaiana, Lagarto e Nossa Sra. da Glória:

- Qualidade e desempenho acadêmico.
- Infraestrutura física de ensino, pesquisa e extensão, biblioteca, recursos de informação e comunicação.
- Relação e comunicação com a sociedade.
- Gestão organizacional e desenvolvimento de pessoal.
- Sustentabilidade ambiental e qualidade de vida.

2.2.2. Plano Diretor de Tec. da Informação (PDTIC)

O PDTIC é o documento que estabelece, para o período compreendido entre os exercícios acadêmicos de 2017 a 2020, objetivos a alcançar e riscos a serem mitigados pelo Núcleo de Tecnologia da Informação (NTI), alinhados ao planejamento estratégico da instituição. No que se refere ao PDTIC da UFS, o PDA encontra-se consoante aos objetivos estratégicos abaixo:

- Prover uma solução integrada em sistemas de informação.
- Implantar e manter uma política de segurança da informação.
- Definir uma política de utilização de recursos de TIC.

2.3. Objetivo geral

Promover uma política de abertura de dados da Universidade Federal de Sergipe, além de mais transparência e publicidade das informações institucionais, fomentar maior participação da comunidade, e interesse de outros interessados em analisar e evidenciar situações de interesse público.

2.4. Objetivos específicos

- Respeitar os dispositivos legais elencados em 2.1, ao passo que foram criados para melhoria do processo de transparência pública.
- Catalogar e reunir dados sobre a instituição em um único portal de fácil acesso por toda a comunidade acadêmica.
- Disponibilizar conjuntos de dados de maneira periódica e consistente.
- Fomentar o desenvolvimento de soluções de TIC sedimentadas nas informações publicadas.

- Promover transparência de informações sobre pesquisa e extensão, facilitando a interação interdisciplinar de departamentos e programas.

2.5. Projeto SmartUFS

O projeto do Portal de Dados Abertos da UFS foi inicialmente idealizado para estar alinhado também com o projeto SmartUFS, idealizado no PROCC/DCOMP, de grande impacto institucional, e que permitirá a UFS administrar seus recursos de maneira inteligente. O SmartUFS é inicialmente um projeto para a Universidade Federal de Sergipe (UFS), mas que tem grande potencial de expansão para outras organizações.

O SmartUFS é um projeto que visa utilizar tecnologia da informação, sensores e dispositivos inteligentes (IoT - Internet of Things) para captar informações a fim de otimizar o uso de recursos disponíveis para a instituição. Para esse projeto, será utilizado a plataforma de cidades inteligentes FIWARE, que possibilita a conexão da ferramenta CKAN de publicação de dados abertos, utilizada como base para o portal de dados abertos da instituição, ao seu portfólio de aplicativos.

Portanto todos os dados captados pelos sensores que não contenham informações de caráter restrito, devem ser publicados no portal futuramente.

3. Critérios de Abertura

Esta seção determina a estratégia de abertura dos dados, levando-se em consideração quatro características: facilidade de acesso à informação, necessidade de cumprimento dos dispositivos legais, e relevância da publicação para a comunidade, consistência da informação. Serão feitas contínuas iterações num ciclo de planejamento, execução, verificação e atuação (PDCA), levando em consideração os itens abaixo.

3.1. Facilidade de Acesso à informação

Serão priorizadas as informações que estiverem inseridas em bases de dados sob tutela do NTI. Em iterações futuras do processo de abertura, devem também ser analisadas as informações em bases de dados de outras unidades, informações em arquivos de texto, e por fim informações impressas em papel que possam ser digitalizadas.

3.2. Necessidade de cumprimento dos dispositivos legais

Em relação a este item, não existe por parte do governo federal nenhuma exigência quanto à natureza dos dados a serem publicados. Entretanto, a confecção deste documento segue os dispositivos legais elencados em 2.1, e as orientações contidas no **Manual Para elaboração de Plano de Dados Abertos**¹⁴, publicado e mantido pelo Ministério do Planejamento Orçamento e Gestão.

3.3. Relevância da publicação para a comunidade

Consoante com as estratégias de combate à corrupção e crimes organizados, a universidade se predispõe a publicar sempre informações relevante no tocante ao interesse da comunidade, como também os conjuntos de dados relacionados necessários à boa compreensão das informações.

3.4. Consistência da informação

A fim de reduzir o risco de suspeitas sobre as informações e atividades exercidas pela instituição, a UFS se reserva a publicar apenas as informações que estiverem consistentes e maduras o suficiente para boa compreensão, comprometendo-se com a constante melhoria da qualidade dos seus dados, visando as futuras publicações.

14

http://www.planejamento.gov.br/secretarias/upload/arquivo/governo-aberto/copy_of_manual_elaboracao_plano_dados_abertos.pdf/@download/file/manual_elaboracao_plano_dados_abertos.pdf

4. Estratégias e Premissas para Abertura

A abertura dos dados leva em consideração estratégias e premissas elencadas nesta seção.

4.1. Estratégias

1. O Levantamento do conjunto de dados candidato à abertura.
2. Priorização e seleção dos dados que serão abertos.
3. Definição de responsáveis pelo preparo e atualização dos dados e detalhamento de plano de ação com metas e prazos.
4. Consolidação da matriz de responsabilidades e definição da governança e do fluxo de aprovação do PDA e revisões.
5. Utilização de metodologia de abertura de dados a ser seguida pelas áreas responsáveis (padrões INDA e INDE).
6. Capacitação dos responsáveis nas áreas de negócio dos dados selecionados para abertura sobre:
 - a. Processo de publicação de dados abertos.
 - b. Processo de catalogação dos metadados no dados.gov.br.
 - c. Processo de catalogação dos metadados na INDE, caso georreferenciados.
7. Definição de arquitetura de abertura para cada sistema.
8. Publicização dos dados catalogados, observando-se o uso de URL fixa. Dados hospedados no sítio do MP, por padrão, serão divulgados na URL <http://opendata.smart.ufs.br> e <http://dados.ufs.br>.

4.2. Premissas

Além de estar alinhado aos princípios e diretrizes mencionados anteriormente neste PDA, o processo de abertura dos conjuntos de dados deve considerar as seguintes premissas:

1. Publicar a maior quantidade de dados considerados relevantes para a sociedade, nos formatos possíveis de serem extraídos e informando as eventuais limitações de qualidade dos dados.
2. Não publicar informações que não públicas ou de caráter estritamente pessoal.
3. Sempre que possível, publicar dados e seus metadados conforme estabelecido no Plano de Ação da INDA, que institui que cada conjunto de dados deve conter, no mínimo:
 - a. Nome ou título do conjunto de dados sucinta.
 - b. Palavras-chave (etiquetas).
 - c. Assuntos relacionados do VCGE18 - Vocabulário Controlado do Governo Eletrônico.
 - d. Nome e e-mail do setor responsável pelos dados.
 - e. Periodicidade de atualização.

- f. Escopo temporal (anual, mensal, diário, bimestral etc. exemplo: dados referentes ao censo de 2011, dados de um indicador mensal).
 - g. Escopo geopolítico (por cidade, por estado, por região).
4. No caso de dados Georreferenciados, deve-se levar em conta as normas e padrões da INDE.
 5. Publicar os dados da UFS seguindo os padrões definidos pela e-PING, pela Infraestrutura Nacional de Dados Abertos - INDA e pela Infraestrutura Nacional de Dados Espaciais – INDE e Governo Eletrônico.
 6. Publicar os dados abertos da UFS no Portal de Dados Abertos da UFS, e no Portal Brasileiro de Dados Abertos.

5. Catálogo de Datasets Publicados

A seguir, o catálogo de datasets publicados com seus respectivos atributos.

Empenhos	
Periodicidade de Atualização	Quinzenalmente
Atributo	Tipo do Atributo
ano	Numérico
data	Data
nro_empenho	Numérico
anulado	Booleano (verdadeiro ou falso)
valor	Numérico
saldo	Numérico
id_ug	Numérico
nome_ug	Texto
sigla_ug	Texto
credor	Texto
observacao	Texto

Discentes	
Periodicidade de Atualização	Quinzenalmente
Atributo	Tipo do Atributo
nome	Texto
ano_ingresso	Numérico
matricula	Texto
nivel	Texto
regularidade	Texto
situacao	Texto
turno	Texto
modalidade	Texto
curso	Texto

Docentes	
Periodicidade de Atualização	Quinzenalmente
Atributo	Tipo do Atributo
nome	Texto
id_unidade	Numérico
unidade	Texto
sigla_unidade	Texto
cargo	Texto

Departamentos	
Periodicidade de Atualização	Quinzenalmente
Atributo	Tipo do Atributo
nome	Texto
sigla	Texto

Cursos	
Periodicidade de Atualização	Quinzenalmente
Atributo	Tipo do Atributo
id	Numérico
nome	Texto
nivel	Texto
turno	Texto
modalidade	texto
unidade	Texto
sigla_unidade	Texto
situacao	Texto

Turmas	
Periodicidade de Atualização	Quinzenalmente
Atributo	Tipo do Atributo
id	Numérico
situacao	Texto
ano	Numérico
periodo	Numérico
id_componente_curricular	Numérico
codigo_disciplina	Texto
nome	Texto
carga_horaria	Horas

Telefones	
Periodicidade de Atualização	Quinzenalmente
Atributo	Tipo do Atributo
id_unidade	Numérico
nome	Texto
sigla	Texto
descricao	Texto
numero	Texto

Restaurantes	
Periodicidade de Atualização	Quinzenalmente
Atributo	Tipo do Atributo
id	Numérico
descricao	Texto

Cardápio	
Periodicidade de Atualização	Quinzenalmente
Atributo	Tipo do Atributo
dia	Data
tipo_refeicao	Texto
nome	Texto
grupo_alimentar	Texto

5. Catalogação no Portal Brasileiro de Dados Abertos

Além da publicação no Portal de Dados Abertos da UFS, os conjuntos de dados da instituição serão catalogados também no Portal Brasileiro de Dados Abertos. A execução desses processos é do NTI, e será realizada preferencialmente por meio de processo automatizado. Porém, caso o conjunto de dados não esteja ao alcance do NTI, a informação deve ser publicada pelo setor responsável por manter a informação, em momento oportuno.

7. Sustentação

A Comissão de Tecnologia da Informação (CTIn) fica responsável por avaliar periodicamente se as premissas determinadas no PDA estão sendo cumpridas e, caso não estejam, determinar ações que viabilizem o devido cumprimento. A comunidade também pode se manifestar, em caso de desarmonia do PDA com os dados do Portal, através dos canais oficiais do Portal de Dados Abertos.

a. Governança

Além dos membros da CTIn, outros atores externos estão habilitados a viabilizar a consistência e periodicidade das informações no portal. A saber:

Ator	Responsabilidade
PROCC	Promoção dos hackathons da UFS para incentivo de acesso ao Portal de Dados Abertos da UFS.
NTI	Disponibilizar a infraestrutura tecnológica necessária ao cumprimento das atividades relacionadas ao desenvolvimento do projeto Portal de Dados Abertos.
Auditoria Interna (AUDINT)	Acompanhar o cumprimento do PDA, manter a política de dados abertos e garantir a colaboração das outras unidades da UFS com o projeto de abertura dos dados da instituição.

8. Monitoramento e controle

A Comissão de Tecnologia da Informação (CTIn), exercerá uma tarefa de avaliação periódica dos dados e rotinas de extração e publicação, com o intuito de melhorar a qualidade das informações extraídas, e otimizar o processo de extração (automatizar processos se necessário). Além disso, são necessárias tarefas de divulgação do portal para a comunidade, como seminários, palestras e hackathons.

a. Melhoria da qualidade dos dados

Periodicamente, a CTIn deve observar os seguintes aspectos das informações publicadas:

- Consistência e veracidade da informação – garantir que a comunidade pode confiar nos dados publicados e que eles serão úteis e consistentes para análise e estudo.
- Nível de abertura – formatos como JSON, GEO-JSON, CSV, ODS, XLSX, XLS são preferíveis a formatos como PDF e formatos de imagem, pois facilitam a leitura das informações por máquina.

b. Comunicação e participação social

A instituição deve promover eventos periódicos com o objetivo de divulgar o portal, a evolução dos dados e o trabalho dos envolvidos no projeto, como também os trabalhos desenvolvidos pela comunidade, utilizando os dados. Dessa forma o portal de dados abertos pode também ser retroalimentado com as novas informações coletadas ou geradas através dos aplicativos desenvolvidos pela comunidade, e se mantém sempre evoluindo.

9. Plano de ação

#	Ação	Atividade	Início	Responsáveis
1	Hackathon UFS (São Cristóvão, Aracaju, e Lagarto)	Eventos para validar a demanda de dados para abertura, e sondar os datasets mais requisitados	2017	PROCC/DCOMP, UFS
2	Desenvolver a arquitetura de software do projeto Portal de Dados Abertos da UFS	Definição da arquitetura do projeto de extração dos dados e publicação no portal	Jan/2018	PROCC/DCOMP
3	Desenvolver o Portal de Dados Abertos da UFS e tecnologias associadas	Estudo da documentação da ferramenta CKAN, e início do processo de instalação e personalização do portal.	Mar/2018	PROCC/DCOMP
4	Estabelecer a governança da política de Dados Abertos	Início de elaboração do PDA	Jul/2018	PROCC/DCOMP
5	Extração dos dados acadêmicos e da estrutura institucional da UFS	Extração dos primeiros datasets. Escolhidos por ordem de facilidade de extração.	Nov/2018	PROCC/DCOMP
6	Extração de dados financeiros (empenhos e liquidações)		Nov/2018	PROCC/DCOMP
7	Aprovar o Plano de Dados Abertos da UFS	Avaliação e aprovação do PDA pelo CTIn	Nov/2018	CTIn
8	Catalogar e publicar os dados abertos da UFS no Portal Brasileiro de Dados Abertos	Catalogação e publicação dos dados abertos da UFS no Portal Brasileiro de Dados Abertos	Mar/2019	NTI
9	Realizar Hackathons para fomentar o acesso ao Portal de Dados Abertos da UFS	Realização de eventos que incentivem o	Jan/2019 em diante	PROCC/DCOMP, NTI, UFS

		acesso ao Portal de Dados Abertos pelos cidadãos e desenvolvedores		
10	Revisão e crítica do processo de extração e publicação dos dados abertos no portal	Revisão crítica do documento do PDA para melhoria da qualidade dos dados e otimização do processo de catalogação e extração	Manutenção contínua	CTIn
11	Reunião para aprovação do novo PDA	Reunião do CTI para aprovação do novo documento	Out – 2020	CTIn

Glossário

Dado – Conjunto de valores que juntos formam informações. Podem ser organizados em formatos diversos, como texto, ou compondo relações entre informações numa estrutura tabular.

Dados abertos – É o dado estruturado acessível pela web num formato que pode ser lido por humanos ou processado por máquinas. Para ser considerado aberto, o dado deve ser disponibilizado sob licença que garanta sua livre distribuição, uso e cruzamento com outras informações.

Dados públicos – São dados mantidos pelo governo que não estão sujeitos a limitações válidas de privacidade, segurança ou privilégios de acesso.¹⁵

Metadados – São informações explicativas sobre o tipo de informação contida num conjunto de dados.

Formato aberto – É uma especificação publicada para armazenar dados digitais, e geralmente mantida por uma organização de padrões não-proprietária e livre de limitações legais no uso.¹⁶

Hackathon – É uma competição entre times com o intuito de utilizar a tecnologia de informação e comunicação para desenvolver idéias e propor soluções para problemas reais.

INDA – Infraestrutura Nacional de Dados Abertos. É um conjunto de padrões, tecnologias, procedimentos e mecanismos de controle necessários para atender às condições de disseminação e compartilhamento de dados e informações públicas no modelo de Dados Abertos, em conformidade com o disposto no e-PING. É a política do governo brasileiro para dados abertos.¹⁷

e-PING – É um padrão de interoperabilidade eletrônica definido pelo governo Federal a fim de viabilizar o trânsito de informações entre órgãos do governo.

FIWARE – É software que serve como plataforma para aplicativos de cidades inteligentes. Foi desenvolvida pela União Européia, tem suporte do projeto FI-PPP (Future Internet Public-Private Partnership), também da União Européia, é aberta e seu intuito é oferecer uma base para o desenvolvimento de aplicações para internet das coisas.

CKAN – É o software de armazenamento e catálogo de dados. É um projeto de código aberto, foi desenvolvido e é mantido pela Open Knowledge Foundation. Foi utilizado como base para o desenvolvimento do Portal de Dados Abertos da UFS.

¹⁵

<http://www2.camara.leg.br/transparencia/dados-abertos/leis-e-principios-dos-dados-governamentais-a-bertos>

¹⁶ https://pt.wikipedia.org/wiki/Formato_aberto

¹⁷

<https://www.governodigital.gov.br/transformacao/cidadania/dados-abertos/inda-infraestrutura-nacional-de-dados-abertos>